LBNL-XXXX

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Measuring Short-term Air Conditioner Demand Reductions for Operations and Settlement

Josh Bode, Michael Sullivan, Joseph H. Eto

January 2012

The work described in this report was funded by the Office of Electricity Delivery and Energy Reliability of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Measuring Short-term Air Conditioner Demand Reductions for Operations and Settlement

Josh Bode, Michael J. Sullivan
Freeman, Sullivan & Co
101 Montgomery Street, 15th Floor
San Francisco, CA

Joseph H. Eto
Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90-4000
Berkeley CA 94720-8136

The work described in this report was funded by the Office of Electricity Delivery and Energy Reliability of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Abstract

Several recent demonstrations and pilots have shown that air conditioner (AC) electric loads can be controlled during the summer cooling season to provide ancillary services and improve the stability and reliability of the electricity grid. A key issue for integration of air conditioner load control into grid operations is how to accurately measure shorter-term (e.g., ten's of minutes to a couple of hours) demand reductions from AC load curtailments for operations and settlement. This report presents a framework for assessing the accuracy of shorter-term AC load control demand reduction measurements. It also compares the accuracy of various alternatives for measuring AC reductions – including methods that rely on regression analysis, load matching and control groups – using feeder data, household data and AC end-use data. A practical approach is recommended for settlement that relies on set of tables, updated annually, with pre-calculated load reduction estimates. The tables allow users to look up the demand reduction per device based on the daily maximum temperature, geographic region and hour of day and simplify the settlement process.

Acknowledgments

This project was sponsored by the California Energy Commission Public Interest Energy Research (PIER) Program under contract #500-05-001. The project team gratefully acknowledges the support of Ron Hofmann, PIER Advisor, and Pedro Gomez, PIER Contract Manager. The team also expresses its appreciation to John Hernandez, Ulric Kwan, Wendy Brummer, Ann George, Melody Augustin, Greg Mandelman and Michael Alexander of PG&E for their contributions to this project. Finally, the project team extends thanks to reviewers who provided helpful comments on the first draft of this report: Steven Braithwait, and Johanna Mathieu.

Table of Contents

List of Figures and Tables

Acronyms and Abbreviations

AC	air conditioning
C&I	commercial and industrial
CAEC	Christensen Associates Energy Consulting
CAISO	California Independent System Operator
CDD	cooling degree-day
CDH	cooling degree-hour
CoA	coefficient of alienation
CPUC	California Public Utilities Commission
CVRMSE	normalized RMSE
DR	demand response
ISO	Independent System Operator
IESO	Independent Electric System Operator
ISO-NE	Independent System Operator – New England
kW	kilowatt
kWh	kilowatt-hour
KEMA	KEMA, Inc.
MAE	mean absolute error
MAPE	mean absolute percentage error
ME	mean error
MPE	mean percentage error
PG&E	Pacific Gas and Electric
RCT	randomized control trial
RMSE	root mean squared error
SCE	Southern California Edison
OPA	Ontario Power Authority
FSC	Freeman, Sullivan Co.
PJM	Pennsylvania-New Jersey-Maryland Independent System Operator
MW	megawatt

Executive Summary

Historically, air conditioner (AC) direct load control programs have been used for emergency operations and to offset the need to build additional peak generation. Their use for day-to-day electric system operations has been limited. However, several recent demonstrations and pilots have shown that AC loads can be used during the summer cooling season to improve the stability and reliability of the electricity grid by providing ancillary services. On average, they noticeably start reducing demand within 60 seconds and typically reach full capability in less than 5 minutes. AC demand and, by connection, demand reduction capability is also typically higher when electricity system needs are higher. Many load reduction control devices also have the ability to shed load automatically when frequency fluctuations are detected and do not require instruction from a central location, providing a safeguard. Importantly, AC loads can be curtailed for short duration periods relatively frequently without affecting customer comfort.

A key issue for grid operations and electricity market settlement is how to accurately measure shorter-term (e.g., ten's of minutes to a couple of hours) demand reductions from AC load curtailments for operations and settlement.[1] Importantly, measurements for settlement and operations need to be conducted quickly (in real time or on a monthly basis), much faster than traditional program evaluations, which are conducted on an annual basis. In addition, measuring demand reductions, sometimes referred to as negawatts, is an entirely different task than measuring power production. While power production is directly measured, by necessity, the measurement of AC curtailments is an indirect estimate. The curtailment is calculated as the difference between electricity use with and without the AC curtailment. However, it is not possible to directly observe what customers would have used in the absence of AC curtailment. To calculate the resource delivered, participant's load patterns in the absence of program participation – the counterfactual, sometimes referred to as the baseline – must be estimated. In doing so, it is important to systematically eliminate or control for alternative explanations for the change in electricity consumption. There are a variety of approaches for measuring the magnitude of AC curtailments with different degrees of complexity, data sources and metering requirements.

This report presents a framework for assessing the accuracy of shorter-term AC load control demand reduction measurements and compares the accuracy of various alternatives for measuring AC reductions using three types of data sources, including feeder data, household data and AC end-use data. The framework essentially tests if the different measurement alternatives correctly calculate demand reductions under different conditions. The study relies on a realistic simulation of AC load curtailments because with a simulation, the real answers are known, making it possible to assess if the measurement is correct and, if not, by how much it deviates from the known AC curtailment.

In total, we tested 10 calculation methods using feeder data, household data and end-use AC data. Each combination of data source and calculation method is considered as a separate measurement alternative. The calculation methods tested include both within- and between

[1] Throughout this report the term "accuracy" refers to both a lack of bias in the measurement and the goodness-of-fit of measurements. For clarity, the metrics to assess accuracy are separated into measures of bias (or lack thereof), goodness-of-fit and variability.

subject estimators. Within-subject estimators use customer's electricity use patterns during days when AC units are not curtailed to estimate AC load absent curtailment operations during actual event days. Between-subject estimators rely on an external control group of AC units that is not curtailed to provide information about electricity use absent curtailment.

While highly accurate results are desirable, there is often a tradeoff between simplicity and incremental accuracy. In order to help gauge the benefit of more complex and costly approaches, each of the measurement alternatives are compared with one of the simplest and least technical approaches – a set of tables with pre-calculated load reduction estimates. The tables allow users to look up the demand reduction per device based on the daily maximum temperature, geographic region and hour of day. They facilitate quick settlement when resources are dispatched and provide operators a quick estimate of the DR resources available for operations.

Key findings from the study include:

- AC impact tables, on average, provide accurate estimates of AC load reductions over multiple events. With a sample of 1,000 accounts, over the course of 15 events, the tables calculate impacts with ±4.5% accuracy with 95% confidence because measurement errors for individual event days cancel each other out. However, for individual curtailments the estimates are less precise.

- Demand reduction measurements for direct load control programs are the least accurate with feeder data and are unreliable for individual curtailment events. Except in feeders with extreme levels of program penetration, AC curtailments tend to be a small share of the feeder load and difficult to accurately distinguish from normal variation in feeder loads – noise. Feeder data includes demand from numerous end-uses and customers that are not enrolled in the load control program. As a result, both the controllable AC load and the demand reductions tend to be a small share of overall feeder demand, making it difficult to accurately identify the demand reductions from the inherent variation in feeder demand levels. Overall, feeder load produces demand reduction estimates that are magnitudes in order less accurate than simple tables.

- Of all 10 calculation methods tested, the 10-in-10 baseline with a 20% in-day adjustment cap is the least accurate. This method is the default approach for estimating impacts for settlement in the California ISO. While it may be adequate for large industrial customers, it does not produce accurate impacts for highly weather sensitive AC load curtailments.

- Regressions produce substantially more accurate AC demand reduction estimates than day-matching or weather-matching baselines, particularly for individual event days. Among baseline methods the weather-adjusted baseline is the most accurate approach for measuring AC demand reductions. However, each of the regression models tested outperformed all of the day or weather-matching baselines regardless of data source. Simply put, regression provides more accurate results than day or weather-matching baselines.

- Because AC control programs typically control a large number of AC units, often well over 100,000 units to be practical, measurements are based on samples. In practice, measurements using smart meter household data produce the most accurate demand reduction results because they can rely on far larger samples of program households. Smart meters lower the cost of using large scale samples and implementing large scale random assignment of curtailment operations. If sample sizes are equal, measurements that rely on directly metered AC end-use data provide the most accurate impacts. However, in practice, sample sizes with smart meter data can be far larger than sample sizes of AC end use data at lower costs.

- With calculation methods that rely on a control group, a simple technique known as difference-in-differences produces more accurate results than simple comparison of means.

The fact that relatively accurate estimates can be obtained using static tables of demand reduction estimates raises several questions. Is it really necessary to use more complex calculations for each individual AC curtailment event? How much value does the incremental accuracy of more complex calculation provide for operations and settlement?

A practical approach is recommended for settlement. It involves using tables with pre-calculated load reductions per AC unit to estimate demand reductions over the summer; conducting a more detailed evaluation at the end of the summer to reconcile settlements and updating the demand reduction tables on an annual basis using a transparent process that allows for independent verification by a third party. As the measurement uncertainty in annual evaluations improves and the number of AC load operations increases, the accuracy of the tables is expected to increase.

While tables with pre-calculated load reductions per AC unit provide accurate estimates of demand reductions over course of the summer, they currently lack the precision needed for grid operations. However, with better measurement in annual AC program evaluations, it may be possible to refine the tables enough to utilize them for operations.

1. Introduction

Historically, air conditioner direct load control programs have been used to offset the need to build additional peak generation. Over the past few years several conceptual and demonstration studies have been conducted that identified and tested the potential of using air conditioner (AC) load control to help improve the stability and reliability of the electricity grid by responding quickly to stabilize the grid.[2345] In addition, several utilities have explored expanding AC load control operations to alleviate transmission constraints or as a way to defer or avoid transmission and distribution equipment upgrades. AC loads can provide the ability to quickly recover from system shocks such as transmission of generation forced outages by removing demand for power from the system. Though less tested, in theory, AC loads can also be configured to increase and decrease in order to regulate the balance of the electric grid. They also can aid the ability of operators to adjust to an unexpectedly fast ramping of system loads.

The demonstration studies confirmed that:

- AC loads can be curtailed quickly and often times respond more quickly than generation. AC units begin to noticeably shut down or cycle compressors within 60 seconds of when the load control signal is sent out and ramp up to the full load reduction capability within less than six minutes.
- Short term AC curtailments have a negligible effect on customer comfort.
- Curtailments can be observed on near real time basis using samples and that the curtailments observed in the sample are matched by fluctuation in feeder loads.

In part because of these demonstration studies, electricity markets have begun to develop products to allow AC load control programs to participate in ancillary services; and internal utility operations teams have expanded exploring how AC load control can be used in grid operations. However, several issues need to be addressed to fully integrate AC load control into electricity markets and operations that have traditionally been reserved for generators.
These include:

- How to accurately measure the demand reductions provided by AC load curtailments;
- How to forecast and bid the available resources;
- How to observe and confirm AC load reduction activation in near real-time;

---

[2] Eto, J., Nelson-Hoffman, C. Torres, S. Hirth, B. Yinger, J. Kueck, B. Kirby (Oak Ridge National Laboratory), C. Bernier, R. Wright, A. Barat, D. Watson. 2007. *Demand Response Spinning Reserve Demonstration.* Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-62761.

[3] Eto, J., C. Bernier, P. Young, D. Sheehan, J. Kueck, B. Kirby, J. Nelson-Hoffman, and E. Parker (2009). *Demand Response Spinning Reserve Demonstration — Phase 2 Findings from the Summer of 2008.* Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-2490E.

[4] Gifford, W., S. Bodmann, P. Young, J. Eto, J. Laundergan. 2010. *Customer Impact Evaluation for the 2009 Southern California Edison Participating Load Pilot.* Lawrence Berkeley National Laboratory. LBNL-3550E.

[5] Sullivan, M., J. Bode, P. Mangasarian. 2009. *2009 Pacific Gas and Electric Company SmartAC Ancillary Services Pilot.* Prepared for Pacific Gas and Electric.

- Automating processes for delivering the specific amount of resources requested by operators by optimizing the dispatch strategy; and
- Developing clearly defined dispatch rules that balance extracting value from DR against exhausting it prematurely.

This report focuses on how to accurately measure shorter-term (ten's of minutes to a couple of hours) demand reductions from AC load curtailments for settlement and operations. It also compares the costs for different measurement alternatives and potential data sources for settlement. The measurements of AC curtailments not only need to be accurate, but also need to be produced quickly. Measuring demand reductions, sometimes referred to as negawatts, is an entirely different task than measuring power production. With generation, electricity output is measured directly. In contrast, the measurement of demand reductions is indirect. It is not possible to directly observe what customers would have used in the absence of a demand reduction event. To calculate the resource delivered, participant's load patterns in the absence of program participation – the counterfactual or baseline – must be estimated. In doing so, it is important to systematically eliminate or control for alternative explanations for the change in electricity consumption.

The fact that it is impossible to directly observe what customers would have used in the absence of load control poses a unique challenge for assessing the accuracy of impact estimates. To assess accuracy it is necessary to know the "true" demand reductions that were delivered. If they are known, it is possible to test which combination of estimation method and data sources produces the correct or most accurate answer. To assess accuracy of different alternatives for estimating AC demand reductions, the study relies on a realistic simulation of AC load curtailments on actual feeder, household and AC end-use data. A simulation is useful precisely because the real answers are known, making it possible to compare estimates from various alternatives to the simulated demand reduction.

This report presents a framework for assessing accuracy of AC curtailments. It also systematically analyzes how the use of feeder, household data and AC end-use data affects the accuracy of the measurement of AC demand reductions. In addition, it compares the accuracy of 11 methods for estimating AC curtailments, including the standard baseline method adopted by the California Independent System Operator (CAISO). One of the simplest and least technical approaches – a set of tables that provides estimates of the load curtailment based on daily maximum temperature, region and hour of day – is used as a benchmark to assess the extent to which more complex approaches for estimating AC curtailments improve accuracy. This basic approach is compared against eight within-subject and two between-subject estimation methods that rely on random assignment of control operations. Within-subject estimators use customer's electricity use patterns during days when AC units are not curtailed to estimate AC load absent curtailment operations during actual event days. They work because the AC curtailment is introduced on some days and not on others, making it possible to observe behavior with and without the load control in effect. Between-subject estimators rely on an external control group of AC units that are not curtailed to provide information about AC units that were curtailed and would have used electricity if they were not instructed to shed load.

While the analysis of AC measurement accuracy relies on data from California, the findings apply more broadly to the measurement of AC curtailments for operations and settlement in any electricity market and for program evaluation in general.

The report is structured as follows. Section 2 provides the market context and summarizes prior research regarding using AC curtailments to provide spinning reserves and accuracy of different settlement methods for DR. Section 3 presents the framework for assessing accuracy of different options for estimating AC curtailments. The accuracy metrics and results for each of the settlement alternatives tested are presented in Section 4. Section 5 assesses the costs of different settlement options and presents a framework for settlement that relies of pre-calculated tables with load reductions per AC unit under various conditions.

2. Context and Prior Research

Air conditioner (AC) cycling is a resource with significant potential for improving the stability and reliability of the electricity grid. The stability and reliability of electricity supply systems depends critically on the ability to balance supply and demand virtually instantaneously at all times. Delays in balancing supply and demand can lead to frequency and voltage fluctuations that compromise the reliability of the electricity grid, often times across multiple states. In essence, this means that there must always be sufficient supply to meet demand.

Balancing supply and demand requires:

- Sufficiently installed capacity to meet even extreme demand levels;
- Generators or other resources that follow electric loads as they rise and fall throughout the day – known as regulation;
- The ability to quickly recover from system shocks such as transmission of generation forced outages; and
- A transmission and distribution system that can deliver power from where it is produced to where it is used.

On a daily basis, a system must carry enough operating reserves to balance the grid and respond to unanticipated events, including forced outages and load forecast errors and other sudden changes in the demand and supply balance. Operating reserves fall into three basic categories – regulation, spinning reserves and non-spinning (or supplemental) reserves. Regulation reserves are synchronized with the electric grid and respond continuously over very short-time scales (i.e., seconds) to achieve balance voltage frequency and maintain power quality. They are designed to respond to small load variations that take place all the time, but are quite small compared to large disturbances. Spinning reserves are designed to maintain grid stability in response to system shocks such as near instantaneous generation and transmission outages. They are typically required to start injecting power into the grid within a minute of notification and ramp up to deliver the full resource within 10 minutes. Historically, spinning reserves have been provided by running some of the generating units on the system at less than full power (i.e., lower heat rates), even though doing so is less economical. Plants running in this way have reserve productive capacity that can be ramped up relatively quickly if a major generating unit is forced off line.[6] In addition, ramping production of power plants up or down too often can impose additional maintenance costs resulting from damage done by rapid increases and decreases in temperature and pressure. Non-spinning or supplemental reserves also are designed to help stabilize the grid in response to large system shocks and, often times, replace spinning reserves when they are dispatched. These are best thought of as stand-by resources that can be started and synchronized with the electric system with enough lead time, which can range from 10 to 30 minutes. Some markets, like the IESO and ISO-NE, have both 10 and 30 minute non-spinning reserves.

All of the above resources require fast response to either continuously balance the grid or to respond to unexpected changes in the supply and demand balance. Electricity system operators

[6] Generators providing spinning reserves typically monitor the electric grid frequency and automatically trigger if the frequency drops below a pre-specified level.

increasingly need resources that can be activated quickly in order to address fast ramping of electricity loads or fast decrease in intermittent power sources, like wind and solar. Fast response resources such as AC loads provide operators additional flexibility in balancing the grid, if they are incorporated into operations.

Historically, AC load control programs have been used as installed capacity or insurance against extreme system demand. These programs target the relatively few hours that drive the need for additional capacity. They have been used as alternatives for building peaking generator units that are more expensive and are only needed a few hours every few years.

Until recently operating reserves have been supplied exclusively by generating equipment either synchronized with the load or in standby reserve status. System shocks that lead to voltage fluctuations (e.g., forced outages or sudden changes in demand) are relatively infrequent and require relatively short operations to stabilize the system, allow generation production to "catch up," and restore reserves. Prior analysis of the frequency of actual operations and their duration confirms this.[7] For example, a 2006 study by Kirby noted that the grid in the Western U.S. experienced 77 voltage fluctuations greater than 0.1hz. Generators have several constraints for providing spinning reserve such as ramping time, minimum on-time or minimum off-time. Historically, demand response – and in particular AC control – was not used to supply reliability services to the power system.

Many of the loads on the electric grid are connected to processes that store energy; and depending on the amount of storage in these processes, it is possible to delay electricity use for some time without significantly effecting the functioning of the processes. Water pumping, air separation, industrial compressors, rock crushing, air conditioning and water heating are examples of such processes. The storage in these processes makes it possible to strategically schedule electricity demand in a manner that can lead to more efficient use of generating capacity and fuel. Many of these processes can drop electricity demand quickly and frequently if events are relatively short with little or no impact on customers.

AC load control programs may provide a lower cost alternative to spinning and non-spinning reserves required to quickly ramp up in the event of a forced outage of generation or transmission facilities or decline in production from intermittent resources such as wind and solar. Loads with control devices also can respond more quickly than most generation facilities and ramp up to full capacity in usually less than five minutes.

The technology for sensing system conditions, scheduling resources and controlling loads has evolved considerably over the past decade as communications and computing systems have continued to develop throughout the U.S. economy. Several recent demonstration projects have shown that direct load control and Auto-DR technology can respond to curtailment requests very quickly, with little impact on customers, and can be integrated with the other resource management systems to provide real time feedback to operators concerning the results of operations.

[7] Kirby, B. 2006. *Demand Response for Power System Reliability: FAQ.* Oak Ridge National Laboratory. ORNL/TM-2006/565.

2.1 Differences Between AC Cycling and Generation

Generation and DR based on AC cycling have different operating characteristics and the differences in these characteristics pose opportunities to use system resources more efficiently as well as problems in calculating the economic value of the resource that is being supplied.

The first and most obvious difference between the operating characteristics of generators and AC load control is that, like all other forms of DR, it is basically the opposite of generation. Instead of injecting power and voltage into the system, it typically removes demand for power from the system.[8] It balances the shortages in supply by reducing demand rather than by increasing supply. For short duration operation, such as operations for spinning reserves, AC loads can be fully curtailed, providing substantially larger reductions than traditional AC control operations. Load curtailment uses no fuel, produces no residual environmental impacts (i.e., greenhouse gasses) and causes no wear and tear on generation, transmission or distribution equipment.

Another important difference between AC cycling and generation is that the load impacts from AC cycling arise from interrupting or curtailing loads for a relatively large number of devices, while the number of generators injecting energy into the electric system is relatively small. Most generators are sized to produce relatively large blocks of power from a single unit (e.g., a 50 MW combustion turbine generator). To produce a load reduction of this magnitude using AC cycling requires controlling between 25,000 and 100,000 AC units depending on the weather. Because of the large number of devices required to produce the load reduction, the likelihood of the total failure of the AC cycling resource is inherently lower than that for a single generating unit or a small number of such units. Just as it is unlikely that 100% or even 50% of generators will not be able to operate when dispatched, it is highly unlikely that most AC load devices will not work when they are dispatched.

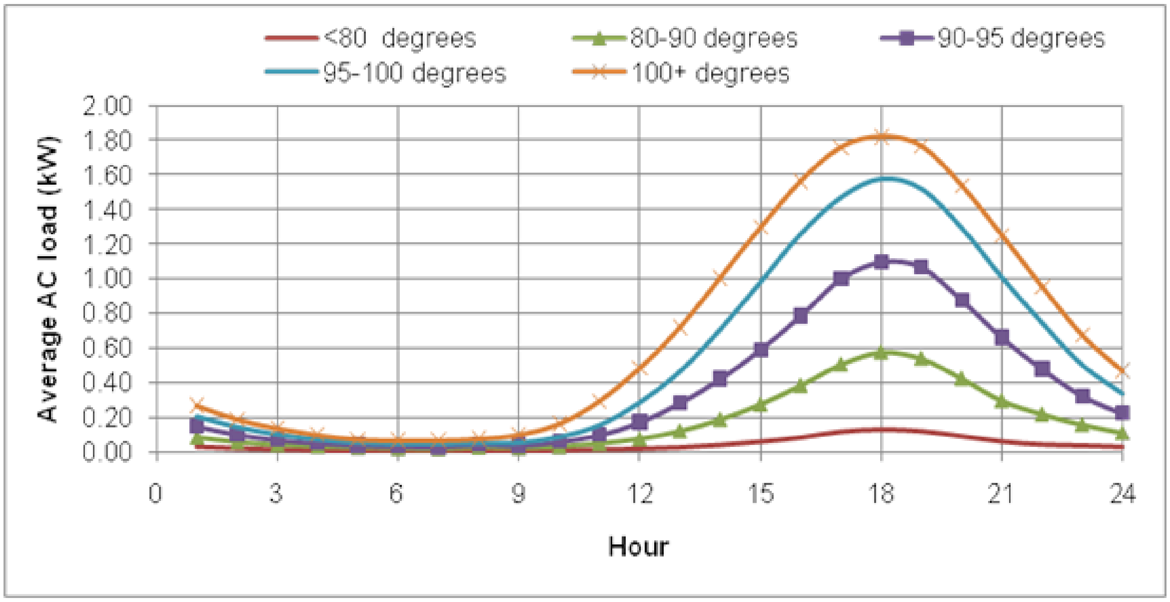
Another key difference is that the output capacity of a generator is known and available throughout the year. If a generator starts and is working properly it will produce a known and highly predictable amount of power regardless of system conditions. This is not true of AC cycling. AC load varies substantially with weather and time of day. While the load reduction is delivered via technology, the AC load itself results from customer behaviors in response to ambient temperature conditions. When temperatures are low, controlling AC loads leads to no or small demand reductions because there is no load to control. On the other hand, when temperatures are high, the same program can produce as much as 100 MW of load reduction in as little as 2 minutes and sustain the reduction for multiple hours, if needed. However, AC load control can supply operating reserves under a limited set of conditions. The sensitivity of AC loads to weather creates challenges in predicting the amount of AC load and load reductions available for grid operations and settlement. Fortunately, while AC load is volatile, it is also predictable from ambient temperature and occurs in conjunction with the times during which installed capacity, operating reserves and ramp speed are most likely to be needed (i.e., when system loads are high due to AC load). In a sense, AC cycling is at the same time the problem and the solution to it.

---

[8] Several electricity systems have recently been concerned not only with shortages in supply but with excess supply. In theory, load control can be used to both inject and remove demand from the system.

Figure 2-1 illustrates the sensitivity of the AC load to weather conditions and is based on the 2009 sample of 550 AC units located in Northern California.

Figure 2-1: Hourly Average Air Conditioner Load for Residential SmartAC Customers by Daily Maximum Temperature



While the area lacks high levels of humidity, the area has a highly diverse climate. For example, in a day where peak temperatures in the Bay Area range between 70 and 80°F, temperatures in the Central Valley will exceed 100°F and often times 110°F. Importantly, average AC hourly demand is almost twice as high in a day with a maximum temperature between 90-95°F than on a day with a maximum temperature between 80-90°F. On a day that exceeds 100°F, the AC load is, on average, three times as high as the average days with a peak between 80 and 90°F. Even within relatively narrow temperature ranges, AC load varies substantially.

Table 2-1 shows the distribution of AC load measurements by daily maximum temperature and hour of day. For the same hour and in the same temperature band, electricity use can be as much as 40% lower or 40% higher than the average for that temperature band.

Table 2-1: Distribution of Average Residential SmartAC Air Conditioner Loads Within Daily Maximum Temperature Bands

Daily Maximum Temperature Band	Hour	Days	Mean	Percentile						
				5th	10th	25th	50th	75th	90th	95th
85°F to 90°F	2:00-3:00 PM	26	0.45	0.24	0.28	0.31	0.47	0.55	0.61	0.65
	3:00-4:00 PM	26	0.59	0.35	0.39	0.45	0.62	0.73	0.77	0.77
	4:00-5:00 PM	26	0.71	0.40	0.49	0.59	0.73	0.84	0.94	0.94
	5:00-6:00 PM	26	0.77	0.48	0.57	0.65	0.80	0.90	0.98	1.00
90°F to 95°F	2:00-3:00 PM	32	0.65	0.35	0.41	0.48	0.68	0.77	0.89	0.92
	3:00-4:00 PM	32	0.84	0.53	0.58	0.63	0.87	0.95	1.03	1.25
	4:00-5:00 PM	32	1.01	0.66	0.70	0.85	1.05	1.13	1.21	1.40
	5:00-6:00 PM	32	1.09	0.79	0.85	0.93	1.09	1.23	1.33	1.43
95°F to 100°F	2:00-3:00 PM	20	0.87	0.55	0.60	0.72	0.87	0.95	1.20	1.23
	3:00-4:00 PM	20	1.10	0.73	0.84	0.97	1.11	1.19	1.45	1.48
	4:00-5:00 PM	20	1.30	0.90	1.04	1.23	1.32	1.36	1.63	1.65
	5:00-6:00 PM	20	1.38	1.04	1.17	1.30	1.41	1.49	1.56	1.63

2.2 Prior Research On Using Air Conditioner Control for Spinning and Non-spinning Reserves

To date there have been several studies that have tested the potential of controlling AC loads in order to provide operating reserves and assessed the ability of integrating control of AC loads into operations. The conceptual framework and the policy reasons for using AC as spinning reserves were detailed in a series of reports by the Oakridge and Lawrence Berkeley National Laboratories.[9, 10] In addition, Lawrence Berkeley National Laboratory, Pacific Gas & Electric (PG&E) and Southern California Edison (SCE) sponsored a series of demonstration studies testing the ability to use AC load control to provide operating reserves. Combined, the studies show that:

- *AC load control reduces demand quickly.* The PG&E 2009 Pilot tested both the ramp time and latency of the load control signal. Because telecommunications or radio signals are used, a lag exists between when the load response signal is sent, when it is received and how long it takes for all units to receive it. In the 71 tests conducted, on average, AC

[9] Kueck, J., B. Kirby, R. Staunton, J. Eto, C. Marnay, C. Goldman, C.A. Martinez. 2001. *Load As a Reliability Resource in Restructured Electricity Markets.* Prepared for the U.S Department of Energy. http://certs.lbl.gov/pdf/load-reliability.pdf

[10] Kirby, B. 2003. *Spinning Reserve From Responsive Loads.* Oak Ridge National Laboratory. ORNL/TM-2003/19.

units begin to noticeably shut down or cycle compressors within 60 seconds of when the load control signal is sent out and consistently reached full capacity within 6 minutes. In addition, to being dispatchable at localized levels, most newer load reduction control devices, including those at PG&E, also have the ability to shed load automatically when frequency fluctuations are detected and do not require instruction from a central location. The frequency set point at which the AC units are automatically shut down is also adjustable. In other words, the load control devices have a built in failsafe mechanism to respond to grid disturbances.

- *<u>The effect of short-duration AC curtailments on customer comfort is negligible.</u>* In 2009, the SCE demonstration measured the extent to which fully curtailing AC loads affected internal building temperatures. The study quantified the increase in indoor temperatures as a function of outside temperatures and the duration of the curtailment operation. It concluded that the vast majority of residents were not aware that they had received an event and 80% did nothing in response to the event. The events had a minimal effect on the temperature of the participating households, with temperature increases of only a few degrees at most. That same year, the PG&E demonstration instructed roughly 2,000 control devices to completely shut down AC compressor electricity use 71 times for 15 minutes at a time. PG&E assessed the effect on customer comfort and the ability of customers to perceive the events by surveying customers whose AC units were instructed to shut off and a control group of customers whose AC units were not curtailed. The two groups did not report a distinguishable difference in comfort or in the ability to perceive AC control operations.
- *<u>AC load drops can be observed on near real time basis using samples</u>*. Each of the studies sampled a subset of the AC units so they transmitted data on electricity use within a minute or less. The data from these individual sites were aggregated and posted on a website so both the AC loads and the demand reductions could be observed in near real time. In addition, users had the ability to view the AC loads and demand reductions in aggregate or for specific feeders.
- *<u>The demand reductions observed in the samples were also observed in the distribution feeder circuits</u>***.** In addition to near real time data on AC loads, the studied intentionally focused on feeders were a large share of the residences had enrolled in the AC load control program. This was done to confirm to system operators that the reductions observed in the sample also could be observed on the feeder loads. The interface was set up so the users could observe in near real time the share of AC units that were on, the AC loads from the sample, the aggregate demand reduction estimate based on the sample and the feeder loads. During hotter days, the drops in AC load could be clearly observed in the loads and matched the demand reduction estimates produced using the sample of AC units.

After testing the capability of AC loads to provide operating reserves, SCE has since been bidding a small share of its roughly 600 MW of AC load resources into the California System Operator (CAISO). This phased approach has been undertaken to fully address the key remaining questions around integration of AC load control into electricity markets and operations. This report focuses on one of the key aspects, namely, how to accurately measure demand reductions from AC load curtailments for settlement and operations.

2.3 Prior Research On Demand Response Settlement Accuracy

Large C&I loads have participated in ISO markets for well over a decade. As a result, much of prior research on estimating demand reductions for settlement have focused on large commercial and industrial customers that have little weather sensitivity compared to AC and residential loads. There also has been some research on different settlement methods for programs that provide customers a rebate if they cut back energy use during high system load hours (Peak Time Rebates). However, none of these studies have assessed different measurement alternatives for market settlement of AC load curtailment programs.

To estimate the counterfactual for settlements, electricity markets have traditionally used day-matching baselines, which are calculated using the participant's load during days preceding the event. With this approach, demand reductions are calculated as the difference between the estimated baseline and actual electricity use patterns during curtailment events. While there are more accurate methods to estimate the counterfactual – or load in the absence of demand response – baselines are useful because they allow settlement to be conducted quickly and are relatively intuitive and easy to understand. Many options exist for calculating baselines. Day matching baselines vary based on:

- *The set of days used to calculate the baseline:* For example, the baseline may be calculated using the 7 days with the highest load out of the 10 weekdays preceding the event or the last 10 weekdays, excluding event days;
- *Application of in-day adjustments:* These adjustments effectively calibrate the baseline up or down based on a comparison between actual loads and baseline estimates during a set of hours preceding the event; and
- *Use of adjustment caps*: When in-day adjustments are used, the rules often limit the magnitude of the baseline in-day adjustments.

Importantly, the common use of day-matching baselines is an outgrowth of the fact that ISO DR products typically target large C&I customers with limited weather sensitivity. The accuracy of many day-matching baselines for settlement in electricity markets or with DR aggregators has been studied in a number of previous studies.

KEMA (2003)[11] compared the accuracy of 6 settlement baselines in 2003 using 646 accounts from multiple regions across the U.S. In total, 206 of the study sites were participants in a DR program, while 440 were not. Baseline error was assessed for non-participants. The non-participant baseline estimates were compared with actual loads for each hour of simulated curtailment periods. For DR participants, the study compared how well the day-matching baselines tested aligned with regression models.

Quantum Consulting (2004)[12] estimated the accuracy of 4 settlement baselines using data from 450 accounts in California, none of which were enrolled in DR programs. The study compared

[11] KEMA, Inc. 2003. *Protocol Development For Demand Response Calculation—Findings and Recommendations.* Prepared for the California Energy Commission. http://www.energy.ca.gov/reports/2003-03-10_400-02-017F.PDF

[12] Quantum Consulting and Summit Blue Consulting. 2004. *Working Group 2 Demand Response Program Evaluation – Program Year 2004 Final Report*. Prepared for the California Public Utilities Commission.

baseline predicted loads to actual loads on three types of days – high load days, low load days and consecutive days. It selected 3 to 7 proxy event days based on system load, with the number of event days selected varied by utility.

Lawrence Berkeley National Lab (2008)[13] also compared accuracy of 7 alternate settlement baselines using data from 32 sites in California, all of which were enrolled in an Auto-DR program. The study compared baseline predicted loads to actual loads on 60 days per site. The proxy event days were selected based on weather. It was the first study to assess accuracy by comparing actual and predicted baseline load for DR participants. All prior studies had drawn conclusions based on either non-participants or comparisons of one estimate to another estimate.

As part of the 2009 California Statewide Evaluation of Aggregrator Programs, Christensen Associates Energy Consulting (CAEC) analyzed the accuracy of the baselines used for settlement.[14] At the time, the program only included large C&I customers with over 100 kW of demand. The included a comparison of baseline calculated impacts to regression calculated impacts for actual events and for pseudo events where the actual baseline was known.

In 2010, FSC analyzed the accuracy of 48 different baseline day-matching methods for a large C&I contractual DR program administered by the Ontario Power Authority, DR-3.[15] The study differed in that it was the first to explicitly distinguish between baseline error and errors in calculating demand reduction attributable to the program – impact error. It also explicitly quantified the extent to which the baseline error was magnified in estimated impacts when the true percent demand reductions were smaller.

In 2010, CAEC studied the issue of baseline accuracy for California's Capacity and Demand Bidding Programs.[16] The focus was on the accuracy of baselines for high volatility and weather sensitive customers and the effect of increasing the same-day adjustment cap. The 2010 study simulated impacts on event-like days. This allows the researchers to know the true demand reduction amounts and assess accuracy by comparing the estimated demand reduction to the simulated reduction.

---

[13]Coughlin K., M.A. Piette, C. Goldman, and S. Kiliccote. 2008. *Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California*. Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-63728. http://eetd.lbl.gov/ea/EMS/EMS_pubs.html

[14] Braithwait, S., D. Armstrong. 2009. *Load Impact Evaluation of California Statewide Aggregator Demand Response Programs. Volume 2: Baseline Analysis of AMP Aggregator Demand Response Program.* Prepared for the California Demand Response Measurement & Evalution Committee.

[15] Bode, Perry and Morgan (2010). Assessment of Settlement Baseline Methods for Ontario Power Authority's Commercial & Industrial Event Based Demand Response Programs. Prepared for Ontario Power Authority. Toronto, ON. November 2010.

[16] Christensen Associates Energy Consulting. 2010. *Highly Volatile-Load Customer Study.* Prepared for Southern California Edison, Pacific Gas and Electric, and San Diego Gas and Electric.

In 2011, KEMA analyzed baseline methods for PJM.[17] The focus was on the use of day matching baselines for large C&I customers and the study tested a number of estimation methods, including day matching and regression, and tested a number of rules for selecting the data for developing the baselines. It also tested each of the settlement baseline methods that were used in different jurisdictions.

As part of the 2009 SCE and PG&E pilots testing the ability of providing operating reserves using AC load control, the demand reductions were calculated. Because of the short duration of events, FSC estimated the PG&E demand reductions using an autoregressive regression model and AC end use data. The models estimated what usage would have been as a function of the load before and after the curtailment, excluding the 15 minutes immediately after the curtailment ended, weather, and binary variables to capture time of day effects. For the first phase of the SCE demonstration, KEMA relied on a simple regression that used the feeder loads in the 10 minutes before curtailment to estimate the counterfactual and demand reductions. However, when the pilot expanded to include additional feeders, the algorithm did not perform reliably during times when the overall load trend was changing, such as in the late afternoon or early evenings. It also was less reliable for the feeders when the share of customers with controllable AC units was lower than in the initial phase. For Phase 2 of the SCE demonstration project, KEMA developed a load matching technique to select patterns of five-minute loads from days without curtailments that were “closest” to loads on the days with curtailments. The approach worked for the feeders with the highest levels of air conditioner control penetration but was not able to detect statistically significant impacts on feeders when the program penetration was lower or during cooler days.

Another key topic in the literature regarding settlement baselines is distinguishing true variation in the performance of DR resources from variation due to measurement error. This topic was recently studied in the context of the automated response of large C&I facilities to dynamic prices.[18] The main conclusion was that measurement error from baselines accounted for a substantial share of the reported event-to-event variability in large customer curtailments.

This body of studies developed a growing consensus on how to assess accuracy of baselines. In specific, the use of proxy event days using DR participant pre-enrollment data is more prevalent in the latter studies because it allows a comparison of estimated impacts to actual known impacts. In addition, the studies drew the conclusion that calibrating baselines using the hours prior to the event improved baseline accuracy substantially. Finally, the majority of the studies indicated that a 10-in-10 day-matching with same day adjustments is the most accurate baseline for large C&I customers.

Based in part on this body of research, the California ISO adopted a 10-in-10 baseline with same day adjustment, capped at 20%, as the standard settlement method for its DR ancillary service

---

[17] KEMA, Inc. 2011. *PJM Empirical Analysis of Demand Response Baseline Methods.* Prepared for the PJM Markets Implementation Committee. http://pjm.com/markets-and-operations/demand-response/~/media/markets-ops/dsr/pjm-analysis-of-dr-baseline-methods-full-report.ashx

[18] Mathieu, J., D.S. Callaway, and S. Kiliccote. 2011. *Variability in Automated Responses of Commercial Buildings and Industrial Facilities to Dynamic Electricity Prices.* Energy and Buildings. *Volume 43, Issue 12, December 2011*

market products. However, the accuracy of these baselines has not been systematically tested for AC curtailments or residential loads until this study.

Except for the 2010 study conducted for OPA, each of the studies focused on the accuracy of the settlement baselines, not on the accuracy of the demand reduction estimates. Effectively, the accuracy of baselines are used as a proxy for the accuracy of demand reduction estimates. The benefit of focusing on the accuracy of baselines is that it makes for a simpler and clearer analysis.

Focusing on baseline accuracy rather than the accuracy of the demand reductions has key shortcomings, however:

- *<u>It is not an assessment of the accuracy of demand reduction measurements.</u>* Good baselines explain much of the variation in electricity use, reducing noise and allowing for better detection of the signal of interest – the demand reduction. However, reducing the noise substantially does not mean that the signal is detected correctly. The ability to do so depends in part on the strength of the signal in comparison to the remaining noise. As we document later, it is possible to explain well over 99% of the variation in feeder loads, but if the signal is relatively small, it is still difficult to correctly detect. The focus on baseline error is analogous to assessing which method is better at reducing noise. While useful, it is not an assessment of how well the signal – the demand reduction – can be detected.
- *<u>Reporting the magnitude of the error relative to the baseline can create the impression the demand reduction measurements are more accurate than they are in reality</u>*. This is particularly true for day matching methods and is best illustrated through an example. Assume the true demand reduction and the true counterfactual – the electricity use absent curtailment – are 2 MW and 10 MW, respectively. A 5% upward bias in the settlement baseline will produce a baseline of 10.5 MW and a calculated load reduction of 2.5 MW (10.5 MW minus the metered load of 8 MW). While the baseline upward bias is 5%, the estimated demand reduction is biased upward by 25%; it is 2.5 MW rather than the actual 2.0 MW. Few readers will take the extra step to calculate how error in the baselines is magnified in the demand reductions estimate.

Appendix A provides a more detailed discussion on why the accuracy of baselines differs from the accuracy of the demand reduction measurements and includes additional examples.

3. Framework for Assessing Accuracy of Demand Reductions Measurements

To assess accuracy of different alternatives for estimating AC demand reductions, we rely on a realistic simulation of AC load curtailments on actual feeder, household and AC end-use data. *In order to assess accuracy, it is necessary to know the actual answer.* If the "true" answer is known, it is possible to assess how close different measurement alternatives were to the right answer. A simulation is useful precisely because the real answers are known.

Rather than focus on the accuracy of baselines – that is, the estimate of what customers would have used absent curtailment – we focus on the accuracy of the demand reductions estimates. Many prior studies have used the accuracy of baselines as a proxy for the accuracy of demand reduction estimates. However, the baselines are simply a means to produce demand reductions estimates. As noted earlier (and in Appendix A), an emphasis on baseline errors is analogous to assessing which method is better at reducing noise. While useful, it is not a direct assessment of how accurately the signal – the demand reduction – is measured.

Figure 3-1 summarizes the general framework used for assessing the accuracy of the demand reduction measurements. To implement the assessment framework, we:

- <u>*Calculated the magnitude of controllable AC loads.*</u> For settlement alternatives that rely on feeder and household data, the strength of the demand reduction signal is weaker than it is with AC end-use data. In this step, we estimate how much of the load from the data source can be controlled for each date and hour. For settlement alternatives that rely on AC end use data, this is an unnecessary step.

- <u>*Selected proxy curtailment events.*</u> In total, 15 curtailment days were randomly selected in 2009 and 2010 from the set of weekdays where daily maximum temperature exceeded 85°F. The threshold was used because in Northern California, AC units are not used much at lower temperatures due to the dry climate and the substantial cooling that occurs overnight. This temperature threshold is exceeded quite frequently during summer months in the inland areas of California where most of the AC load control program participants are located. The curtailment event start times were randomly selected between 12 PM and 10 PM with durations of one or two hours. The proxy curtailment events are designed to reflect use of AC load control during spinning or non-spinning operations or to provide fast ramping capability.

- <u>*Simulated the demand reductions*</u>: The simulated demand reductions rely on the variation observed in historical percent reductions of air conditioner load from annual impact evaluations. The percent reductions incorporate the effect of weather plus a random variation component. The size of the reduction is conservative because it relies on results from normal, multi-hour control operations, when AC loads are not fully shed.[19] The percent reductions were then applied to the controllable AC loads to produce the simulated demand reduction. Importantly, with this process, the demand reductions for

[19] Because of the short duration of spinning and non-spinning operations, it is possible to fully shed AC loads, producing larger demand reductions and a stronger signal. Because the evaluation percent reductions are themselves estimates, they contain some degree of measurement error for individual events. As a result, the process applied likely overstates the true variation in percent load reductions. In both cases, this places a more stringent test on the ability of the measurement alternatives to detect the demand reductions since it requires detecting weaker signals with more volatility.

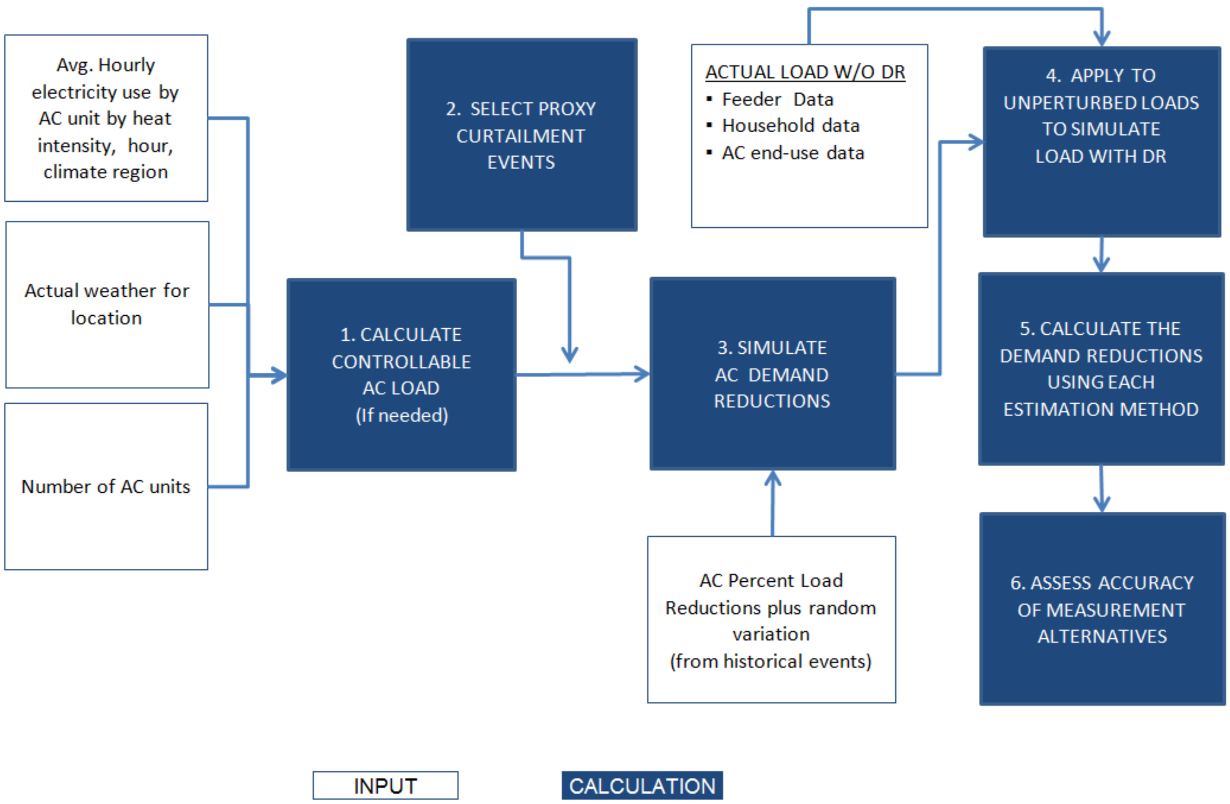
each curtailment event are known, making it possible to test how accurately each of the different settlement alternatives measures the load drop.

- *Applied the demand reductions to unperturbed loads.* During each of the proxy curtailment event periods, the simulated demand reductions were subtracted from the unperturbed loads. In other words, we knew the actual demand with and without the simulated curtailments as well as the magnitude of the demand reductions. The demand reductions per AC unit were similar for the feeder, household and AC end use data simulations. However, the reductions were a smaller percent of the overall feeder and household loads than of the AC loads. Households had additional non-AC electric loads that reduce the signal to noise ratio. And, feeder loads had additional noise from households that did not enroll in the load control program, including customers without AC units and commercial and industrial loads.

- *Calculated the demand reductions using each data source and 10 estimation methods:* The demand reductions were calculated using the feeder, household and AC end use data. The calculation methods tested are detailed later in this section, but included day and weather matching methods, regressions and approaches that relied on control groups. Each combination of a data source and calculation method was considered a separate measurement alternative.

- *Assessed the accuracy of each of the settlement alternatives:* For each of the curtailment events, we knew the true load patterns without curtailment and the true demand reductions. In other words, we had the answer key and could grade which measurement approaches got the right answer, and, if not, how close they were to doing so. As a result, we were able to assess the accuracy of each measurement alternative. To standardize the comparison, we used metrics designed to assess if the measurement alternatives systematically over or under-reported demand reductions (bias) and metrics that summarized how close the measurements were to the true demand reductions (goodness-of-fit). These metrics are detailed later in this section.

Figure 3-1: Overview of Framework



The remainder of this section provides details regarding each step, including information about the underlying sources used and, where appropriate, simplified examples. The section concludes with discussion on how sampling error was incorporated into the estimates of measurement accuracy. Sampling error and sample sizes have different implications for within-subject estimators – that is, estimates that rely on using data from days without curtailment – than they do for between-subject estimators – that is, calculation methods that rely on a comparison of two groups. As a supplement, Appendix B provides additional detail on the feeder, household and AC end-use data sources.

3.1 Calculation of Controllable AC Loads

The controllable AC loads were calculated by multiplying hourly AC load profiles by the appropriate number of AC units.[20] For example, if the maximum temperature for a particular day was between 90°F to 95°F in the California Central Valley and the time was between 3 and 4 PM, based on historical data, the average AC unit uses 0.92 kW. In addition, for each feeder, the number of controllable AC units is known. Suppose the feeder had 88 AC units (this was actually the average number of controllable AC units per feeder). The controllable AC load for

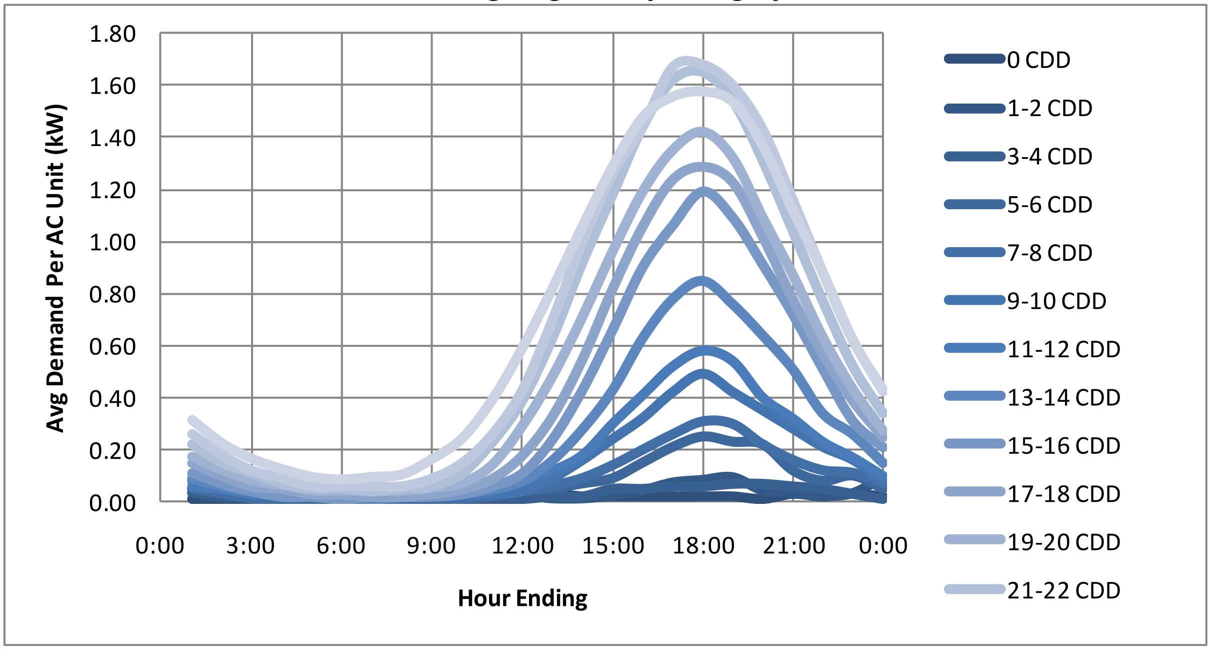
[20] This step was not necessary for AC end use data.

the feeder is estimated by multiplying the demand per AC unit (0.92) by the number of controllable AC units (88). In the example, there is 81.1 kW of controllable load.

In 2009, PG&E collected data on AC end use electricity data for a sample of 547 AC units enrolled in the load control program. The sample was stratified across three different climate regions.[21] Importantly, AC loads for these units were not curtailed during the time period. In other words, the sample reflected how customers enrolled in SmartAC naturally operated their AC units without the intervention of load control. This AC load data was used to create 24 hourly load profiles, based on temperature conditions, for each of three geographic regions.[22] The temperature conditions were defined based on the cooling degree days and ranged from 0 to 22 in increments of two.

As an example, Figure 3-2 shows the average demand per AC units for each hour and heat intensity bins for the Very Hot climate region. Appendix C includes tables with the average AC use for each climate region, daily temperature profile and hour.

Figure 3-2: Average Demand per AC Unit in a Very Hot Region (Fresno/Bakersfield) By Cooling Degree Day Category and Hour



[21] The coastal region, which includes San Francisco and Oakland, was excluded since less than 1% of SmartAC residential participants are located in that region

[22] Cooling degree days are a measure of heat intensity and are designed to reflect the conditions under which space cooling is needed. It is calculated by subtracting a base value, 65˚F, from the average temperature in the day. Unlike a metric like daily maximum temperature, it is not only the peak but overnight heat build-up which can significantly affect AC use.

3.2 Selection of Proxy Event Days and Hours

In total, 15 weekdays per year were randomly selected from all days when the SmartAC average customer maximum temperature met or exceeded 85°F during the months of July through September. Days with a daily maximum temperature below 85°F generally have little AC load in California because overnight temperatures cool off substantially more than in humid regions. Overall, SmartAC customers are generally in inland areas that frequently exceed this threshold in the summer months of June through September. For example, in the hottest region – the Fresno/Bakersfield region of the Central Valley –daily maximum temperatures exceeded 90°F on 76% of summer days. In the second hottest region – the Northern part of the Central Valley near Stockton and Sacramento – daily maximum temperatures exceeded 90°F on 43% of the days.

For each event day, event start times were randomly selected between 12 PM and 10 PM. Since system shocks and the need for ramping speed can occur at various times, it was important to introduce variation in the timing of the proxy curtailments. In addition, the duration of events was randomized between one and two hour durations. While in practice most spinning and non-spinning dispatches are very short, lasting less than 15 minutes, the event durations were at the hourly level for 4 reasons. First, hourly events allowed comparison across feeder, household and AC data. This was necessary because most of the historical household data was at hourly increments. Second, while long events are rare, reserves are required to be able to deliver resources for at least two hours. This means that the baseline settlement method needs to be robust for such events. Third, one or two hour events provide additional information about the accuracy of different measurement for grid operations other than spinning and non-spinning reserves. The fourth reason is that AC loads can also provide operators flexibility during multi-hour ramp events.

Table 3-1: Characteristics of Test Events

Date	Start	End	Duration (hours)	Daily Maximum Temperature (˚F)			
				Very Hot (Fresno / Bakersfield)	Hot (Sacramento / Stockton)	Warm (Diablo Valley / South Bay)	Population Weighted Average
July 13, 2009	1:00 PM	2:00 PM	1.0	95.9	93.8	94.1	94.3
July 15, 2009	2:00 PM	4:00 PM	2.0	105.9	99.3	91.1	97.0
July 16, 2009	4:00 PM	6:00 PM	2.0	106.5	99.4	92.6	98.2
July 27, 2009	8:00 PM	9:00 PM	1.0	105.9	100.6	90.3	97.6
August 13, 2009	1:00 PM	2:00 PM	1.0	97.8	93.2	86.5	91.4
August 18, 2009	5:00 PM	6:00 PM	1.0	100.9	94.8	85.3	92.5
August 21, 2009	10:00 PM	11:00 PM	1.0	100.3	98.3	95.3	97.3
August 26, 2009	1:00 PM	3:00 PM	2.0	99.8	91.9	84.3	90.7
August 28, 2009	12:00 PM	1:00 PM	1.0	98.7	96.8	99.7	98.1
September 2, 2009	2:00 PM	4:00 PM	2.0	101.3	96.1	96.9	97.3
September 3, 2009	9:00 PM	11:00 PM	2.0	101.0	99.0	94.9	97.2
September 11, 2009	9:00 PM	11:00 PM	2.0	97.1	97.9	95.7	96.4
September 17, 2009	6:00 PM	7:00 PM	1.0	93.4	90.1	89.1	90.2
September 18, 2009	9:00 PM	10:00 PM	1.0	100.3	96.1	96.0	96.6

Date	Start	End	Duration (hours)	Daily Maximum Temperature (˚F)			
				Very Hot (Fresno / Bakersfield)	Hot (Sacramento / Stockton)	Warm (Diablo Valley / South Bay)	Population Weighted Average
September 23, 2009	6:00 PM	7:00 PM	1.0	100.3	97.2	88.5	94.0
July 14, 2010	12:00 PM	2:00 PM	2.0	99.5	94.0	88.6	93.0
July 16, 2010	8:00 PM	9:00 PM	1.0	105.1	99.4	88.5	96.1
July 19, 2010	4:00 PM	6:00 PM	2.0	104.1	96.9	82.1	91.8
July 23, 2010	12:00 PM	1:00 PM	1.0	100.7	92.8	84.0	90.6
August 3, 2010	6:00 PM	8:00 PM	2.0	100.2	93.4	85.0	91.6
August 16, 2010	8:00 PM	9:00 PM	1.0	98.5	93.7	83.1	90.4
August 23, 2010	4:00 PM	6:00 PM	2.0	96.2	94.5	96.4	95.5
August 26, 2010	7:00 PM	8:00 PM	1.0	104.7	91.0	81.3	90.2
September 1, 2010	8:00 PM	9:00 PM	1.0	94.7	92.9	93.9	93.5
September 2, 2010	1:00 PM	3:00 PM	2.0	102.1	97.9	96.8	98.2
September 3, 2010	9:00 PM	11:00 PM	2.0	103.2	97.2	90.9	95.8
September 6, 2010	2:00 PM	4:00 PM	2.0	94.8	91.3	93.5	92.8
September 28, 2010	12:00 PM	2:00 PM	2.0	100.6	99.7	100.4	100.0
September 29, 2010	2:00 PM	4:00 PM	2.0	100.0	100.0	96.4	98.1
September 30, 2010	10:00 PM	11:00 PM	1.0	99.8	92.8	86.9	91.6
2009	**12:00 PM**	**11:00 PM**	**1.4**	**100.3**	**96.3**	**92.0**	**95.3**
2010	**12:00 PM**	**11:00 PM**	**1.6**	**100.3**	**95.2**	**89.8**	**93.9**

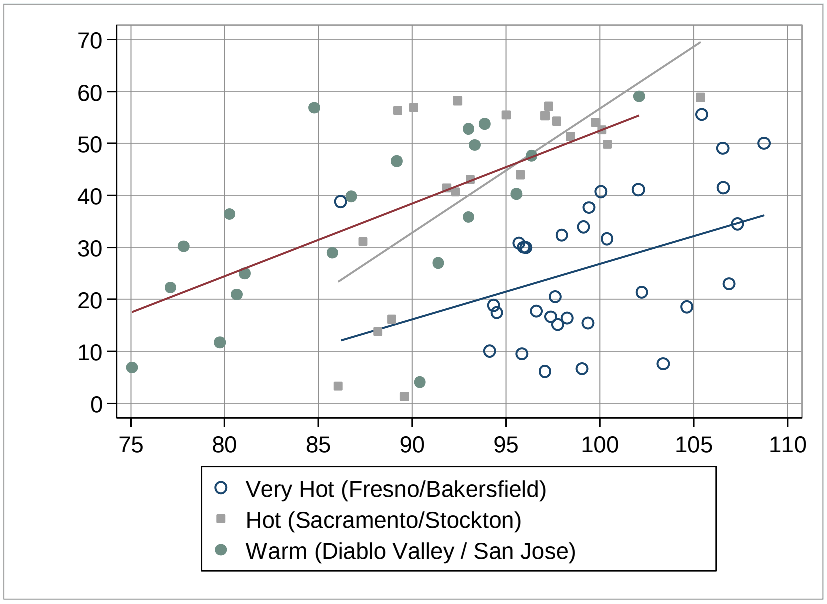
Table 3-1 summarizes each of the proxy events including date, start and end time, durations and temperature conditions. It includes both daily maximum temperature values for each climate region and the daily maximum temperature weighted for the mix of SmartAC residential participants across the PG&E service territory. The temperatures during the proxy curtailment periods were generally substantially lower than the daily maximum because the events ranged from 12 PM to 11 PM. While event days were hotter than average, they were neither extreme nor rare. On average, event period temperatures were about 87°F, but ranged from a low of 71°F to a high of 98°F due to variation in weather and event periods. During the cooler days, AC units may have been in operation in areas with less heat intensity, such as the Inland portions of the Bay Area.

3.3 Simulation of Demand Reductions

The simulated demand reductions were calculated by multiplying the controllable load by a percent load reduction. For each proxy curtailment event, the percent load reductions were based on regression analysis of historical impacts by temperature plus a random variation component. The feeder data example from above assumes a 35% reduction leads to a 21.1 kW drop in AC demand (81.1 kW x 35%). On a per AC unit basis, the simulated demand reduction in the example is 0.32 kW (0.92 x 35%).

Figure 3-3 plots the estimated percent load reductions and temperature conditions for the average AC unit in each of the three warmer PG&E climate regions.

Figure 3-3: Percent Load Reductions by Temperature Conditions for the Average AC Unit (SmartAC Measurement and Evaluation Sample)



Between 2008 and 2010 PG&E called a total of 34 events on a representative sample of AC units to evaluate program impacts and assess how impacts varied under different event conditions. While the program is designed to operate when the system is peaking, many of the evaluation test events were called when temperatures were cooler than what would be expected for normal operations in order to understand the demand reduction potential under different weather conditions.. For most of the events, AC units were controlled for four hours under a partial control strategy - 50% cycling – which still allows the AC unit to cool homes. A few of the events lasted up to six hours or employed more intense cycling strategies such as 66% cycling, but were generally limited to half the sample and were excluded from the historical data on percent load reductions.

Because the simulated impacts are based on a 50% cycling strategy, they yield smaller impacts that instructing devices to shut off the AC compressor (load shed operations). In other words, the strength of the signal – demand reductions – shrinks relative to the back ground noise when the AC units are not fully curtailed. In all cases, the accuracy of the impact estimates improves with load shedding because it produces a stronger signal that is easier to distinguish from background noise.

3.4 Application of Demand Reductions to Unperturbed Loads

Next, the demand reductions were subtracted from the actual, unperturbed electric load data, at the unit of analysis, for each relevant hour. As a result of this process, the true loads with and without the AC curtailment as well as the true demand reductions were known. In the feeder data example, this involves subtracting the aggregate reduction in AC use, 21.1 kW, from the total feeder demand, 7,159 kW. The aggregate air conditioner reduction is very small compared to the feeder load for two reasons. First, only a small fraction of customers on the example feeder were enrolled. Second, the example curtailment event day is a relatively mild day in the Central Valley. Although the daily maximum temperature was assumed to be between 90°F and 95°F, there is little humidity in the area and temperatures cool substantially overnight. If instead the 21.1 kW were substracted from the household data of SmartAC participants on the feeder, 150.1 kW, the percent reduction is a substantially larger share of the relevant load.

3.5 Demand Reduction Calculation Methods Used

The next step was to estimate the demand reductions using different measurement alternatives. Each combination of the calculation method and data source is considered as a distinct measurement alternative because both affect the accuracy of measurements. The different data sources have different amounts of non-AC end uses and "background noise" from which the demand reduction – the signal – must be detected.

This was done using feeder, household and AC end use data. Table 3-2 summarizes the calculation methods used. A total of 10 different demand reduction calculation methods were applied to feeder, household and AC end-use data. The least technical approach – a set of tables that provides estimates of the load curtailment based on daily maximum temperature, region and hour of day – is used as a benchmark to assess the extent to which more complex AC curtailment measurement alternatives improve accuracy.[23] The 10 calculation methods can be classified into 2 broad categories: within and between subject estimators.

[23] The table amounts to a deemed estimate of the demand reduction based on the hour of day, temperature and location. While it provides quick estimates of load reductions for operations and settlement, it is not a calculation method per se. It still must be updated periodically based on more sophisticated evaluation methods.

Table 3-2: Summary of Measurement Alternatives Tested for Accuracy

Type of Estimator	Method	Calculation	Data Source			Summary Description
			Feeder	House Data	AC end-use	
Within subject estimators	Day matching baseline	10-in-10 with a 20% in-day adjustment cap	X	X	X	A subset of weekdays when units were not cycled is identified and the average is calculated for each hour to produce a baseline. The days are selected from the 10 non-event weekdays closest to the load curtailment day. The baseline is calibrated or adjusted using information about demand patterns in the hours preceding the curtailment (in-day adjustment). Demand reductions are calculated as the difference between the adjusted baseline and metered load. The process for weather matching baselines is similar except that the baseline load profile is selected from non-event days with similar daily maximum temperature and then calibrated with an in-day adjustment.
		10-in-10 without an in-day adjustment cap	X	X	X	
		Top 3 in 10 without an in-day adjustment cap	X	X	X	
	Weather matching baseline	Profile selected based on daily maximum temperature without an in-day adjustment cap	X	X	X	
	Regression	Treatment variables and no day or hourly lags or leads	X	X	X	Regression analysis quantifies how different, observable factors such as weather, hour of day, day of week, location and load curtailments affect AC electricity use patterns. With regressions, the impacts are usually directly estimated through the model parameters that reflect the effect of load control operations – known as treatment variables. With treatment variables, the impacts are the difference between the regression estimates of AC use with and without load control. Regression models can be informed by electricity use patterns in the day prior (day lags) and in the hours before or after an event (lags or leads).
		Treatment variables with a day lag	X	X	X	
		Treatment variables with hourly lags and leads	X	X	X	
		No treatment variables but use of hourly lags and leads	X	X	X	
Between subject estimators	Random assignment of load control operations	Comparison of Means		X	X	AC load control program participants are randomly assigned to groups that do and do not have their AC unit instructed to reduce or shed AC load. Any differences between the two groups are random, not systematic. The group that is not subject to the load curtailment is typically referred to as the control group and provides information about normal electricity use patterns in the absence of AC curtailment. Impacts are calculated as the difference in average demand between the group that is and is not dispatched (comparison of means). The estimate can be refined by assessing inherent differences between the two groups in hot non-event days and netting them out of the demand reduction calculation (difference-in-differences).

Within-subject estimators use customer's electricity use patterns during days when AC units are not curtailed to estimate AC load absent curtailment operations during actual event days. They include demand reduction calculation methods such as individual customer regressions and day and weather matching baselines. They work because the AC curtailment is introduced on some days and not on others, making it possible to observe behavior with and without the load control in effect.

Between-subject estimators rely on an external control group of AC units that are not curtailed to provide information about AC units that were curtailed and would have used electricity if they were not instructed to shed load. While there are several types of between-subject calculation methods we only consider two simple options that rely on random assignment to load control operations: a simple comparison of means and a difference-in-differences calculation.

3.5.1 Impact Estimate Tables

Impact estimate tables are the least technical approach and are typically constructed at the AC unit level. They are essentially a detailed lookup table. They rely on actual AC electricity use patterns and historical percent load reductions under specific, discrete temperature conditions and can be customized for specific regions. For example, based on historical data from the AC end use sample that was not subjected to AC load curtailments, the average electricity use for a Central Valley residential AC unit is 1.2 kW between 3 PM and 4 PM when the daily maximum temperature is between 95°F and 100°F. Historically, when cycled at those temperatures, AC use dropped by 35%, on average, though there was variation in the percent reductions estimated for individual events. Based on the decision matrix, we would estimate that impacts for a similar day are 0.42 kW per AC unit (1.2 kW x 35%). In practice, the actual unperturbed load without DR for that day and the percent load reduction may be different than those in the table, leading to some error. For example, if the true uncontrolled AC load were 1.1 kW and true percent impacts were 30% (0.33 kW), the table impact estimates would be too high by 22%. While not complex, the approach is practical and low cost. It serves as a useful baseline for assessing how much value is added by using more complex baseline calculation approaches.

To assess the accuracy of impact estimate tables it was necessary to replicate the fact that a sample was used to create the table. In this process, a random sample was drawn from the population and used to construct the tables. The tables produced using the sample provided an estimate for the demand reductions in the population. The accuracy of the tables was assessed by gauging how close the predictions from the tables were to the known, simulated demand reductions in the population.

3.5.2 Day and Weather Matching Baseline Methods

Day-matching is a less technical approach than regressions for developing an estimate of what electricity use would have been in the absence of load control. It also has been widely used by ISO's for settlement of DR products designed for large C&I customers. This calculation method relies solely on electricity use patterns when the AC unit is not controlled. A subset of weekdays when units were not cycled is identified. These are usually days in close proximity to the event day. In each case the electricity use in each hour of the identified days is averaged to produce a baseline. While there are more accurate methods to estimate the counterfactual – or load in the absence of demand response – baselines are useful because they allow settlement to be conducted quickly and are relatively intuitive and easy to understand. Many options exist for calculating baselines. Day-matching baselines vary based on:

- The set of days used to calculate the baseline: For example, the baseline may be calculated using the 7 days with the highest load out of the 10 weekdays preceding the event or the last 10 weekdays, excluding event days;
- Application of in-day adjustments: These adjustments effectively calibrate the baseline up or down based on a comparison between actual loads and baseline estimates during a set of hours preceding the event; and
- Use of adjustment caps: When in-day adjustments are used, the rules often limit the magnitude of the baseline in-day adjustments.

Day-matching baselines are often supplemented with corrections to incorporate information about usage patterns in the hours preceding an event – usually referred to as in-day or same-day adjustments. In-day adjustments are common and reduce the error between the unadjusted baseline and actual loads. Appendix D provides a detailed example of how in-day adjustments are applied.

Many jurisdictions cap the magnitude of in-day adjustments in order to limit the potential for manipulation of the baseline. With AC load control, gaming is not a concern for two reasons. First, AC direct control participants are not directly compensated for performance or non-performance during events and have no incentive to manipulate load in order to change the baselines. Second, AC curtailments typically occur without any advance announcement. If used for grid operations, the short lead times for activations of AC curtailments does not provide enough time to manipulate baselines.

The high weather sensitivity of AC loads has key implications for baseline in-day adjustments, however. As noted earlier, a day with a daily maximum of 95°F can lead to more than twice the electricity use of a day of 90°F. This has two key consequences. First, ratio adjustments often need to be large and can exceed 200%. Second, they are more volatile when applied to directly metered AC end-use because the baseline period can include many days when AC units were not on. Dividing any value by a very small number (e.g., 1.2 kW/0.2 = 12) leads to very large adjustments.

Weather-matching baselines are a variation from day-matching methods. The main difference is that the comparable days are based on average hourly load patterns during non-event days with similar weather conditions, as defined by temperature bins. These days may or may not be immediately prior to the curtailment event. Continuing our earlier example, to produce a baseline for a weekday with a daily maximum temperature between 90°F and 95°F, the first step would be to identify weekdays with similar temperatures and without AC curtailments. Suppose there were six such days. The electricity use for each time period would be averaged for those six days to produce a baseline. As with day-matching baseline, the baseline can be calibrated or adjusted using actual usage patterns in the hours preceding an event. Given the sensitivity of AC load, this is recommended.

3.5.3 Regression Analysis

Regression analysis quantifies how different, observable factors such as weather, hour of day, day of week, location and cycling affect AC electricity use patterns. With regressions, the impacts are directly estimated through the model parameters. In other words, the impacts are the difference between the regression estimates of AC use with and without load control.

The analysis consists of applying regression models separately at the unit of analysis. The regression specification is common over all units but estimated coefficients vary for each unit. The variables in the regression specifications model time-based and weather-based impacts. The fact that each feeder has its own specification automatically accounts for variables that are relatively constant for each unit, such as geographic location, mix of load control switches versus smart thermostats and the strength of the communication network. Because the coefficients are specific to the unit, they can better explain the variation in weather sensitivity and load patterns.

Regression models work because AC load control naturally produces an alternating or repeated treatment design. The primary intervention – AC load controls – is present on some days and not on others, making it possible to observe AC use with and without cycling under similar conditions. A repeated introduction and removal of curtailment events allows for an assessment of whether the outcome – electricity consumption – rises or falls with the presence or absence of AC cycling. This approach works if 1) the effect of AC cycling does not spill over into other days and 2) there is electricity use data for a sufficient number of days that match actual dispatch conditions.

The four regression models tested as part of assessment vary primarily based on whether they predict loads as a function of the loads observed during prior or neighboring time periods (often referred to as auto regressive models). The first model relies solely on external factors to estimate the electricity demand and demand reductions. These include time of day, type of day (weekday/weekend) and total heat intensity over the prior 24 hours. The second model uses the same external factors plus the usage during the same hour of the prior day to estimate usage patterns and demand reductions. The third model uses the same external factors plus information about loads observed in the hours neighboring the curtailment event. In the first three models, the demand reductions are estimated by variables that interact with the curtailment and temperature conditions. The fourth model is the same as the third one, with one key exception. It does not estimate the demand reduction based by using a variable to represent the event. Instead, impacts are calculated by taking the difference between the regression prediction of electricity use absent the curtailment and the actual load, much like day-matching baselines. Appendix E contains the mathematical expression of the regression models tested.

Although regression models can measure demand reduction relatively accurately, they bring added complexities. They are less transparent to non-statisticians and more complex to implement for operations and settlements.

3.5.4 Estimation Methods with Control Groups

Another way to estimate demand reductions is by using a control group of customers that is not curtailed during the event. In essence, the electricity demand patterns by the group that was not curtailed are used to infer what the usage patterns of the curtailment group would have been absent the curtailment.

However, on its own, using a control group does not guarantee more accurate results. A good control group has customers that, on average, look like and behave in the same manner as the customers whose AC loads were controlled except for the curtailment event. To eliminate alternative explanations for differences in electricity use, it is critical that the only systematic difference between the two groups is the fact that one group had their AC units curtailed while the other group did not. To put it differently, if two groups behave almost exactly the same during all hours of the year except for the hours when AC units were curtailed, it is reasonable to conclude the difference in electricity demand is due to the curtailment event.

The best way to ensure there are no systematic differences between the two groups is to randomly assign customers to the curtailment and control groups and use large sample sizes. This approach is known as an experiment or a Randomized Control Trial. It is widely regarded as the best evaluation method. Because of random assignment, on average, both groups can be expected to have similar characteristics such as household size and to experience the same weather, economic conditions and occupancy patterns. That is, random assignment of AC load control events eliminates alternative explanations for changes in demand by eliminating systematic differences between the two groups. The only systematic difference between the two groups is whether or not they were curtailed. However, differences between the two groups can arise due to random variation in sampling. With larger groups, it is less likely that substantive differences exist.

Using control groups with random assignment has several benefits. It ensures the demand reduction estimates are unbiased and precise, provided large samples are used. More over the process is extremely easy to understand and allows the use of calculation methods that are highly transparent and simple to execute. It does not need to rely on a complex mathematical model or rules for selecting match days. In our assessment, the demand reductions were calculated in one of two ways:

- *A simple comparison of means:* With this approach, for each time period, demand reductions are estimated as the difference between the group that did not have their AC loads curtailed and one that did.
- *A weather-matched difference-in-differences calculation:* This approach is useful when sample sizes are smaller. The demand reduction is calculated as the difference between the two groups, but then adjusted with one additional step. We subtract out differences between the two groups during days without curtailments and similar weather. This nets out differences that are irrelevant and mainly due to sampling variation.

Figure 3-4 illustrates an example of calculating AC demand reductions with smart meter data, random assignment to operations and large sample sizes. The example is based on the 2011 SmartAC evaluation, which was still underway when this report was being written. In the evaluation, customers were randomly assigned to 1 of 10 groups. During each of nine test events, one of the randomly assigned groups was operated while the other nine groups acted as a control group. During actual program events, 9 of the 10 randomly assigned groups would be operated. For the example day, a randomly assigned group of 10% of the devices, or 14,000 devices, was activated for curtailment. The remaining 90% or 124,000 devices acted as the control group and provided information about normal electricity use without the curtailment.

Figure 3-4: Example of Using a Control Group to Estimate Demand Reductions

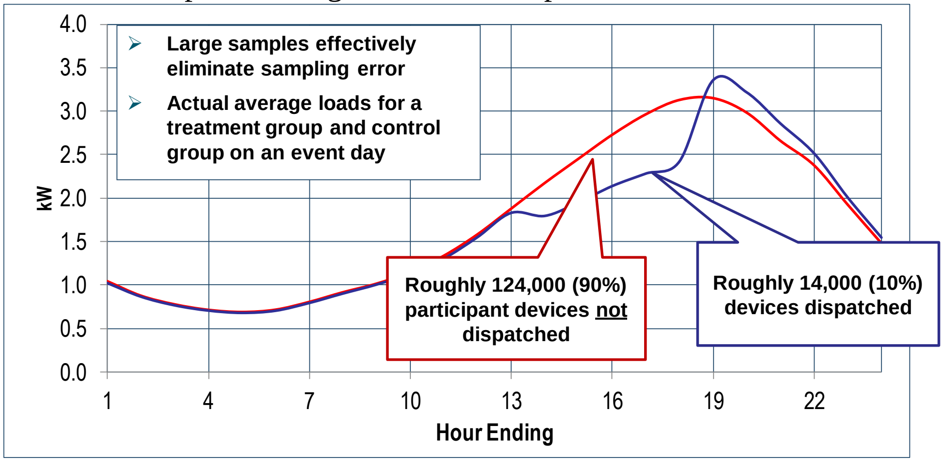


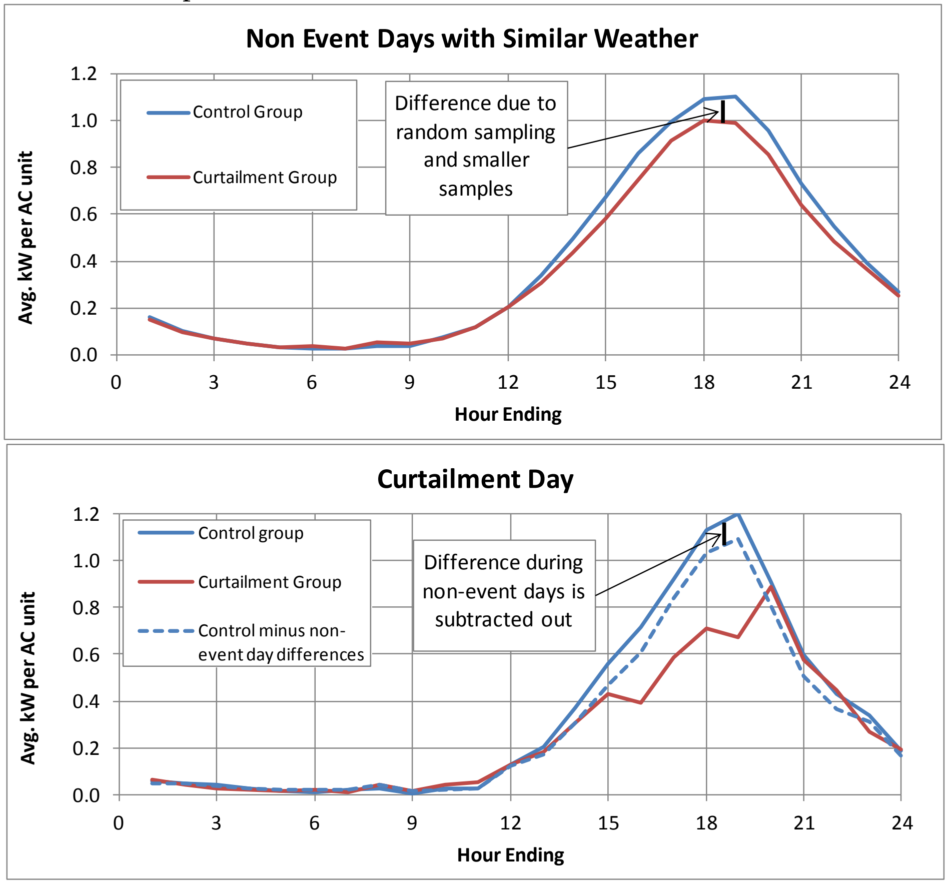
Figure 3-4 shows actual metered loads for the two groups without any adjustments. It illustrates the accuracy and simplicity of the approach. Loads leading up to the event were identical for both groups and clearly diverged during the curtailment period. The large sample sizes effectively eliminated almost all random sampling error, while the randomized assignment ensured both groups were identical except for the load curtailments.

Not all utilities are able to measure AC demand reductions using random assignment with groups of similar magnitude, for reasons discussed below. Even within a larger utility like PG&E, demand reduction measurements are sometimes required at a more localized level, reducing sample sizes. Smaller sample sizes introduce the potential for differences due to random variation in the sampling – that is, it can introduce differences that are unrelated to curtailments. With smaller sample sizes a difference-in-differences calculation can remove these irrelevant discrepancies and improve both the accuracy of the results.

Figure 3-5 illustrates the calculation using AC end-use data. In the example, 600 units were sampled and 300 were randomly assigned to the curtailment group and 300 were assigned to act as a control group.

The two groups have a clear difference in electricity use during non-event days and in the hours leading up to the event in the curtailment day. This difference is entirely the result of random sampling and the smaller sample sizes. Taking the simple difference between the two groups during the curtailment period clearly overestimates the demand reductions. To correct this, the differences observed during the non-event days with similar weather are subtracted out. In the example, the true demand reduction is 0.25 kW per AC unit. Without the adjustment, using the control group produces demand reductions of 0.40 kW. With the adjustment, the demand reduction estimate is 0.30 kW, a clear improvement.

Figure 3-5: Example of Weather-matched Difference-in-Differences Calculation



For many AC load control programs, it is not always feasible to implement randomly assigned control groups with large sample sizes. The first consideration is load control device communications. Newer AC load control programs such as PG&E's SmartAC typically use systems that can transmit cycling instructions to individual AC units. For example, it is possible to instruct the load control device of a house to shed load and to instruct the load control devices at an adjacent house not to do so. Older AC cycling programs such as SCE's often rely on one-way communications where all units in a region respond to a radio signal. This has practical implications. For example, for a program like PG&E's SmartAC, it is possible to randomly assign the participants into 10 groups and withhold a group of over 14,000 accounts from being dispatched during each event, rotating the control group. For a program without load control devices that can be directly addressed, this is not feasible. To create a control group, they would need to install “placebo” or inactive load control devices for a subset of households. The second consideration is costs. With smart meters in place, the costs of deploying and using large samples with several thousand customers to estimate load reductions is dramatically lower. The data also can be retrieved remotely and analyzed within days. Without smart meters in place, there is no such luxury and sample sizes are a legitimate concern. The ability to use extremely large sample sizes and ensure accurate representation of the population of interest are two of the most attractive features of smart meter household data. Utilities that had not yet deployed smart meters would need to install data collection devices, adding a substantial cost per unit included in the samples. As a result, they would likely need to rely on substantially smaller samples of either household AC end-uses or households.

3.6 Metrics for Assessing Accuracy

Accuracy refers to how close a measurement is to the actual value. The settlement alternatives produce estimates of the AC load control impacts. Comparing them to the true impact values allows us to assess the accuracy of each settlement alternative. Since AC load control programs aggregate tens and sometimes hundreds of thousands of the AC units, the focus is less on the accuracy of estimates for individual AC units or feeders and more on the overall accuracy of the results for the program and for larger zones in the electric grid.

Because the demand reductions were simulated, the answer key was available and we could assess how close each of the measurements was to the right answer. To standardize the comparison, we used metrics designed to assess if the measurement alternatives systematically over or under-reported demand reductions (bias) and metrics that summarized how close the measurements were to the true demand reductions (goodness-of-fit). An accurate estimator produces results that are on average unbiased and minimize amount of error for individual periods (i.e., it has a high goodness-of-fit).

Table 3-3 summarizes the metrics for bias and goodness-of-fit used to assess the different measurement alternatives. It includes a brief description and the corresponding mathematical equations.

Metrics of bias indicate whether the demand reduction measurements tend to be disproportionately positive or negative. In other words, they indicate the extent to which the measurement alternatives tend to over or underestimate the true demand reduction. The bias can be reported in an absolute basis by computing the mean error or on a percentage basis by reporting the mean percentage error. In this report, the comparison of bias are made using the mean percentage error (MPE) because it allows for a direct comparison of the results from feeder, household and AC end-use data since the metric is standardized. For example, an MPE of 5% indicates that the measurement alternative, on average, overestimates demand reductions by 5% and an MPE of -10% indicates the baseline on average underestimates demand reductions by that percentage.

A number of goodness-of-fit statistics are summarized. For each of the metrics selected, lower values indicate lower amounts of error (or higher accuracy). In comparing different measurement alternatives, however, normalized metrics such as the mean absolute percentage error, the normalized root mean squared error, and the coefficient of alienation are favored because they allow for comparisons across the different data sources.

Table 3-3: Metrics for Assessing Bias and Goodness-of-Fit

Type of Metric	Metric	Description	Mathematical Expression
Bias	Mean Error (ME)	The mean error (ME) indicates whether on average the errors tend to be disproportionately positive or negative. This metric is presented in the same units as the original and is not normalized for comparison for different types of data.	$ME = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$
	Mean Percentage Error (MPE)	The mean percentage error (MPE) indicates the percentage by which the measurement, on average, tends to over or underestimate the true demand reduction.	$MPE = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}}$
Goodness-of-Fit	Mean Absolute Error (MAE)	The mean absolute error (MAE) reflects the average error, regardless of positive or negative direction, and is in the same units as the original data. It does not weight larger error more so than smaller errors.	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i - y_i\rvert$
	Mean Absolute Percentage Error (MAPE)	The mean absolute percentage error (MAPE) is a measure of the relative magnitude of errors across event days, regardless of positive or negative direction. It is normalized allowing comparison of results across different data sources.	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert$
	Root Mean Squared Error (RMSE)	The root mean squared error (RMSE) is sensitive to larger errors. The squaring process gives disproportionate weight to very large errors, which are then recalibrated by taking the square root. This metric is not normalized and can only be compared between models whose errors are measured in the same units.	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$
	CV(RMSE)	This metric normalizes the RMSE by dividing it by the average of the actual demand reduction.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$
	Pearson's Chi-Squared	This test is applicable to both continuous and discrete data. For each observation, the squared difference between demand reduction measurement and the actual demand reduction (the expected value) is divided by the actual demand reduction. These values are subsequently summed.	$\chi^2 = \sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{y_i}$
	Coefficient of Alienation	The coefficient of alienation measures the amount of variation that is not explained (or accounted for) by the demand reduction estimator. Values closer to zero indicate good explanatory power.	$CoA = \sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{y_i^2}$

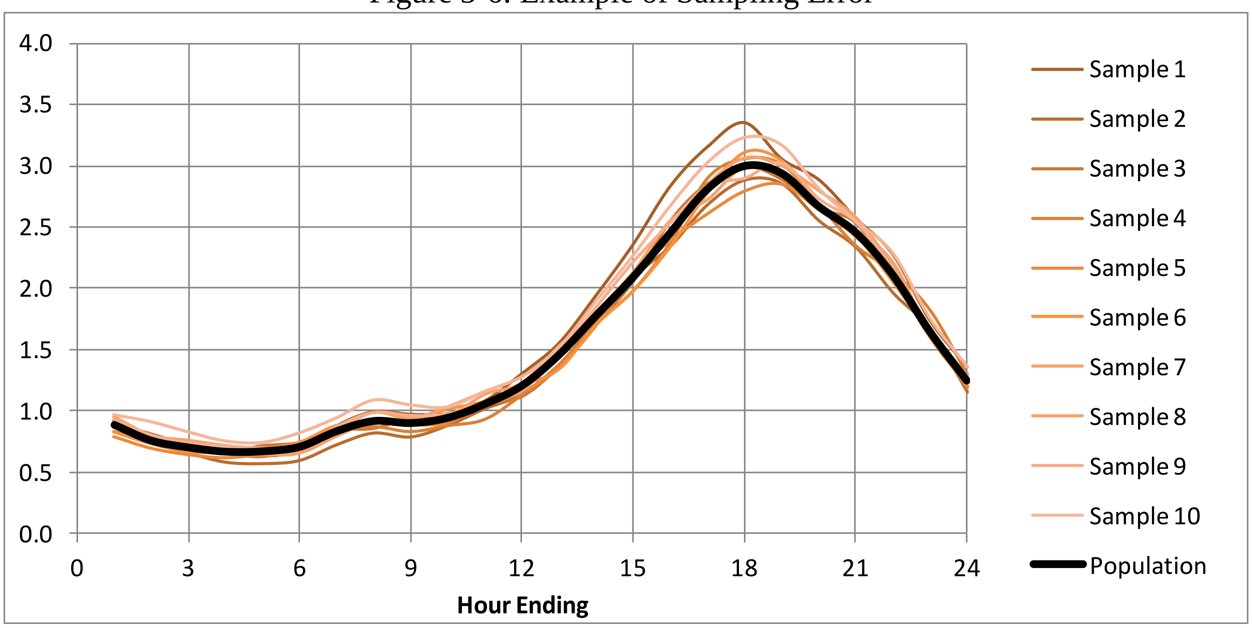
In making that comparison, it is important to understand if the settlement alternative is unbiased on average and accurate for individual curtailment hours. A settlement alternative that produces correct measurements on average can perform poorly for individual events. This occurs if the errors cancel each other out. To better understand this, a simplified example is useful. Assume there are only two event days and the true impacts for each are 0.2 kW and 0.6 kW per household. Two settlement options can both produce the right average answer, 0.4 kW, but perform differently for individual event days. An option that estimates 0.4 kW for the first and second event is clearly inferior to one that accurately estimates the 0.2 kW and 0.6 kW impacts for the two events – even though both estimate the average event impact correctly.

3.7 Implications of Sample Sizes for Measurement Error

AC load control programs aggregate many tens or hundreds of thousands of small resources that are dispersed both across distribution circuit feeders and within them. As a result, metering each unit is impractical due to sheer number of participants and the use of samples is recommended.

The use of samples has implications for the accuracy of the demand reduction measurements. Smaller samples can introduce error from random variation – known as sampling error. Figure 3-6 illustrates sampling variation. It shows the hourly load shape on September 2, 2010 for 10 random samples of 300 households and compares them to the population hourly load shape.[24] Some of the samples reflect the population load shape extremely well, while other load shapes noticeably over or underestimate it.

Figure 3-6: Example of Sampling Error



The random sampling process is expected to reflect the population, if repeated over and over. However, in practice, samples are typically only drawn once. As sample sizes increase, the variation in sample draws decreases and precision increases. With a sample of 300 customers,

[24] Except for sampling error, these random samples are representative of AC load control program participants.

the electric loads used for the analysis were within ±11.4% of the population values 95% of the time. With samples of 500, 1,000 and 2,000 customers, the electric loads were respectively within ±8.8%, ±6.2 and ±4.4% of the population values 95% of the time.[25] The precision of samples does not increase in proportion to increase in sample sizes. For example, doubling the sample size from 500 to 1,000 does not lead to a twofold increase in precision. In general, the first hundred sample points decrease the margin of error the most.

Sampling error can play a larger role when demand reductions estimates rely on control groups than it does when the reductions are calculated based on usage in prior non-curtailment day (within-subject methods). To illustrate, assume the control group sample drawn happens to have +5% sampling error and, for simplicity, that the curtailment group has no sampling error. The true population load absent curtailment for the population happens to be 1.50 kW per AC unit and the true demand reduction is 0.40 kW per AC unit. The control group, however, produces a counterfactual of 1.58 kW and an estimated demand reduction of 0.48 kW (1.58-1.10 kW) instead of the true reduction of 0.40 kW. With the control group, the 5% sampling error in the control group, leads to a 20% error in the demand reduction measurement in the example.

With within-subject methods, the estimate is not affected by a second sample such as a control group. This is because they rely on electricity use patterns during days when the AC loads are not curtailed to estimate the demand reduction. To illustrate with the above example, if the sample of the curtailment group happens to have +5% sampling error, the demand reduction for the sample would be 0.42 kW instead of the population true value of 0.40 kW. But the sampling error would also lead to an upward bias on non-curtailment days of roughly the same amount. In effect, sampling error tends to affect demand reduction estimates less with within-subject methods. Appendix G contains a table showing how the overall margin of error varies as a function of both sampling error and estimation error.

To incorporate the effect of sampling error, the demand reduction estimation process was replicated 100 times using 100 different randomly drawn samples for methods that relied on control groups and for the impact estimation tables. Appendix H provides more detail on the process used to incorporate sampling error.

[25] The margin of error calculation takes into account the variability of individual observation as measured by the coefficient of variation (the average divided by the standard deviation), which can vary across jurisdictions and hours of day. For the data used, the mean over peak period (2 PM to 6 PM) was 2.602 kW and the standard deviation was 2.610, producing a coefficient of variation near 1. The 95% margin of error is calculated by the following formula, where 1.96 is the critical value with a 95% interval and the FPC is the finite population correction.

$$ME = \frac{\left[1.96 \cdot \left(sd/\sqrt{n}\,\right)\right]}{\mu} \cdot FPC$$

4. Accuracy of Measurement Alternatives

In choosing settlement alternatives the tradeoffs between accuracy, transparency, cost and integration into existing systems needs to be understood. For a settlement alternative to merit more extensive consideration, it needs to meet a basic threshold of accuracy. The demand reduction measurement used for settlement need to produce results that do not have a tendency to over or underestimate the demand reduction and that produce precise results.

This section summarizes the accuracy of the measurement alternatives tested. In total, we tested 10 calculation methods using feeder data, household data and end-use AC data. Each combination of data source and calculation method is considered as a separate measurement alternative. The calculation methods tested include both within and between-subject estimators and are detailed in Section 3.5. All of the results are reported on a per AC load control household basis to allow for comparison of the different data sources.

While highly accurate results are desirable, there is often a tradeoff between simplicity and incremental accuracy. In order to help gauge the benefit of more complex and costly approaches, each of the measurement alternatives are compared with one of the simplest and least technical approaches – a set of tables containing pre-calculated demand reduction estimates. The tables allow users to look up the demand reduction per device based on the daily maximum temperature, geographic region and hour of day.

The remainder of this section is structured as follows. First, the results from using simple impact estimate tables are presented to serve as the benchmark. This is followed by the results of the regression, day-matching and weather-matching baselines are presented for each of the data sources – individual AC unit data, aggregated AC data, household data and feeder level data. Finally, the results from approaches that rely on using a control group are presented. These results are presented for individual AC unit and household data. The section concludes with a summary of the key findings.

4.1 Accuracy with Impact Estimate Tables

Impact estimate tables are the least technical approach and serve as the benchmark for assessing more sophisticated methods for estimating AC load control impacts. They are essentially a detailed set of lookup tables with deemed savings. The user looks up the relevant climate region, the temperature conditions for the day and the hour of day for an estimate of the load reduction per AC units; they then multiply the values times the number of units curtailed in each relevant geographic region to estimate the aggregate impacts.

Impact estimate tables are based on actual AC electricity use patterns and historical percent load reductions under specific, discrete temperature conditions, in specific geographic areas. They are a simple way to predict or estimate the demand reduction for settlement and are typically produced on a per AC unit basis so they can be used to estimate demand reductions for localized demand reductions or partial dispatches of the AC load control reduction capability. A sample of the impact estimate tables is included in Appendix F.

For example, based on historical data, the average electricity use for a Central Valley residential AC unit is 1.2 kW between 3 PM and 4 PM when the daily maximum temperature is between 95°F and 100°F. If, historically, AC use drops by 35% under those temperatures, based on the decision matrix, we would estimate that impacts for a similar day are 0.42 kW per AC unit with 50% cycling (1.2 kW x 34%). If 20,000 AC units were controlled in the region, the aggregate reduction would be 8.4 MW.

In practice, the true counterfactual and the true percent load reduction may differ from the estimates produced using the tables. This leads to error for individual curtailment events. However, if the tables are accurate they should produce estimates close to the true values and should lead to accurate payment of settlement over the course of the summer. The accuracy of the tables also depends in part on the magnitude of the sample(s) used to produce the tables and degree of sampler error. Using impact estimate tables is similar to "deemed savings" estimates commonly used for energy efficiency programs. While not complex, the approach is practical and low cost. As a result, it serves as a useful benchmark for assessing how much value is added by using more complex baseline calculation approaches.

Table 4-1 summarizes whether or not the tables over or under estimated the demand reduction for the population. As discussed in Appendix G, the tables and the predictions were estimated 100 times using different sample sizes to reflect the uncertainty from the sampling process. The table shows the distribution across the samples of per household load without load control, the true population impacts, the predicted impacts and the bias and goodness-of-fit metrics introduced in Section 3.6. In each case, the table presents the distribution of these metrics across all 100 sample draws or simulations. In practice, a sample is typically only selected once per year. As a result, the tables indicate the likelihood that a single sample will yield different levels of accuracy.

The tables produce relatively accurate estimates of the true demand reductions over the summer. This is reflected by the mean percent error, a metric for bias, which ranges from -0.1% to -0.4% for the average sample, meaning that it does not over or underestimate the true demand reductions. However, since samples can only be drawn once, the potential for sampling error is largest with smaller sample sizes. As the underlying sample sizes increase, the potential for using a sample that is biased decreases substantially. In particular, with a larger sample size there is less uncertainty regarding unperturbed AC loads and, by connection, more accurate load reduction estimates. With the relatively small sample size of 300 customers, in 90% of the samples drawn the mean percent error for the average event was between -6.7% and 7.8%. To put this in context, if the true impact for the average event were 0.210 kW per household, 90% of the sample draws would produce estimates between 0.196 kW and 0.226 kW per household. This improves substantially with larger sample sizes. With a sample size of 1,000 AC units, 90% of the impact estimates had a mean percent error between -3.7% and 4.0%.

However, as will be detailed later in this section, while the tables do not over or underestimate demand reductions by much over the course of the summer, they are less accurate for individual event days. This is can be seen by the goodness-of-fit metrics from the tables to other measurement alternatives. For an individual event, the tables produce, on average, errors between 30% and 40%, but because the errors are not systematically biased they cancel each other out. The goodness-of-fit statistics do not improve much as sample size increase. This has implications for its potential application for operations and settlement. While a simple table can be used for settlement, a system operator may care if 50 or 60 MW of demand reductions is available for operations.

Table 4-1: Accuracy Metrics for Impact Estimate Tables

Sample Size	Metric		Average	Std. Deviation	Distribution Percentiles				
					5th	25th	Median	75th	95th
100	Impacts	Pop. average true impact	0.21	0.00	0.21	0.21	0.21	0.21	0.21
		Impact estimate	0.21	0.02	0.18	0.19	0.20	0.22	0.24
	Bias	Baseline percent error	2.3%	9.7%	-10.5%	-5.0%	0.6%	8.3%	19.6%
		MPE (Demand Reduction)	-0.1%	9.3%	-12.5%	-7.3%	-1.8%	6.0%	17.0%
	Goodness of Fit	MAE	0.07	0.01	0.06	0.07	0.07	0.08	0.09
		MAPE	36.9%	5.7%	31.0%	32.6%	35.2%	39.4%	48.0%
		CVRMSE	0.43	0.05	0.37	0.39	0.41	0.44	0.52
		Coefficient of alienation	1.22	0.30	0.92	1.03	1.14	1.30	1.82
300	Impacts	Pop. average true impact	0.21	0.00	0.21	0.21	0.21	0.21	0.21
		Impact estimate	0.21	0.01	0.19	0.20	0.21	0.21	0.22
	Bias	Baseline percent error	2.8%	4.2%	-4.0%	0.3%	2.6%	5.4%	10.9%
		MPE (Demand Reduction)	0.4%	4.1%	-6.7%	-2.1%	0.2%	2.7%	7.8%
	Goodness of Fit	MAE	0.07	0.01	0.07	0.07	0.07	0.08	0.08
		MAPE	36.5%	2.8%	32.2%	34.7%	36.2%	38.2%	41.4%
		CVRMSE	0.42	0.02	0.38	0.40	0.41	0.43	0.46
		Coefficient of alienation	1.15	0.13	0.98	1.06	1.13	1.22	1.39
500	Impacts	Pop. average true impact	0.21	0.00	0.21	0.21	0.21	0.21	0.21
		Impact estimate	0.21	0.01	0.19	0.20	0.21	0.21	0.22
	Bias	Baseline percent error	2.5%	3.6%	-3.2%	-0.1%	2.7%	4.6%	9.3%
		MPE (Demand Reduction)	0.1%	3.5%	-5.5%	-2.8%	0.1%	2.3%	6.7%
	Goodness of Fit	MAE	0.07	0.00	0.07	0.07	0.07	0.07	0.08
		MAPE	36.3%	2.1%	33.2%	35.0%	36.1%	37.4%	40.4%
		CVRMSE	0.41	0.02	0.39	0.41	0.41	0.42	0.44
		Coefficient of alienation	1.14	0.08	1.01	1.09	1.14	1.18	1.30
1000	Impacts	Pop. average true impact	0.21	0.00	0.21	0.21	0.21	0.21	0.21
		Impact estimate	0.21	0.01	0.20	0.20	0.20	0.21	0.21
	Bias	Baseline percent error	2.4%	2.4%	-1.5%	0.6%	2.2%	3.9%	6.6%
		MPE (Demand Reduction)	0.0%	2.3%	-3.7%	-1.7%	-0.2%	1.5%	4.0%
	Goodness of Fit	MAE	0.07	0.00	0.07	0.07	0.07	0.07	0.08
		MAPE	36.2%	1.4%	33.8%	35.3%	36.0%	36.8%	38.9%
		CVRMSE	0.41	0.01	0.40	0.41	0.41	0.42	0.43
		Coefficient of alienation	1.13	0.06	1.05	1.10	1.13	1.17	1.25
2000	Impacts	Pop. average true impact	0.21	0.00	0.21	0.21	0.21	0.21	0.21
		Impact estimate	0.21	0.00	0.20	0.20	0.20	0.21	0.21
	Bias	Baseline percent error	2.4%	1.4%	0.1%	1.4%	2.2%	3.5%	4.7%
		MPE (Demand Reduction)	0.0%	1.4%	-2.4%	-0.9%	-0.2%	1.0%	2.3%
	Goodness of Fit	MAE	0.07	0.00	0.07	0.07	0.07	0.07	0.08
		MAPE	36.1%	0.9%	34.9%	35.5%	36.1%	36.6%	37.8%
		CVRMSE	0.41	0.01	0.40	0.41	0.41	0.42	0.43
		Coefficient of alienation	1.13	0.04	1.07	1.10	1.12	1.15	1.20

4.2 Accuracy for Within-subject Estimators

As noted earlier, within-subject estimators use electricity use patterns during days when AC units are not controlled to estimate the counterfactual and demand reduction in days when units are controlled. One set of days for a customer (or group) is used to predict load patterns for another set of days. This is the most widely used approach to estimate impacts for DR programs. Not surprisingly, to date, discussions on settlement of DR in electricity markets have focused almost exclusively on day-matching baselines and other algorithms to establish what electricity use would have been in the absence of the demand reduction. This is an outgrowth of the fact that demand response programs for large C&I customers were the first type of DR integrated into electricity markets. In practice, this class of estimators includes regression models that do not make use of a control group and all day-matching and weather-matching baselines. In total, eight different estimators are analyzed in this section. Three are day-matching baseline, one is a weather-matching baseline and four are regression-based estimates.

Each of these estimators was applied to individual AC end-use data, aggregated AC end-use data, household data and distribution feeder circuit data. Because the underlying data affects accuracy, the results are presented by source.

4.2.1 Individual Air Conditioner End-use Data

With AC end-use data, the percent load reduction – the "signal" – is largest compared to the background noise. There are no other participant end-uses and the data does not include non-participants. As a result, a priori, one would anticipate that all things equal the results of the AC end-use data would be more accurate. However, there are two defining features of AC end-use data that affect the accuracy of the results. First, AC units are highly weather sensitive and quite often not in operation. As a result, the day-matching baselines often require substantial adjustments based on the measured pre-event data. Because the AC use during the pre-event period can be small or zero, the ratio for in-day adjustment can become very volatile. Second, data collection of AC end-use data is a high cost proposition and in practice leads to smaller samples than for other data sources.

Table 4-2 summarizes the results for the average event for each regression, day-matching and weather-matching settlement alternative. It shows the extent to which the measurement options over or underestimates the baseline and, more importantly, the demand reductions. In other words, the table summarizes the degree of bias, if any, for each measurement alternative. It also highlights the fact that baseline and demand reductions are in fact different. The results are presented on a per household basis. In other words, during the average simulated event, the average household used 0.78 kW of AC load and controlling it produced 0.20 kW of load reduction.

Table 4-2: Accuracy by Settlement Alternative for Average Event
Individual Air Conditioner Data

Method	Calculation	Actual load without DR	Predicted load without DR	Mean % Error	Actual Impact	Predicted Impact	Mean % Error
Day Matching	1. 10-in-10 baseline with 20% in-day adjustment cap	0.78	0.51	-35.0%	0.20	-0.07	-135.9%
	2. 10-in-10 baseline with uncapped in-day adjustment	0.78	0.97	24.2%	0.20	0.38	93.7%
	3. Top 3-in-10 baseline with uncapped in-day adjustment	0.78	0.97	24.8%	0.20	0.39	95.9%
Weather Matching	4. Weather baseline with in-day adjustment	0.78	0.89	14.3%	0.20	0.31	55.4%
Regression	5. With treatment variables and lags or leads	0.78	0.77	-1.3%	0.20	0.20	0.8%
	6. With treatment variables and a day lag variable	0.78	0.78	0.5%	0.20	0.20	2.3%
	7. With treatment variables and pre and post event hour variables	0.78	0.79	1.9%	0.20	0.21	6.3%
	8. With no treatment variables and with pre and post event hour variables	0.78	0.77	-0.7%	0.20	0.19	-3.4%

The results for day and weather-matching methods tend to over or underestimate actual demand reductions by a substantial amount. They produce inaccurate estimates of both the baseline and of the impacts. This is due to the nature of AC operations and in-day adjustments.[26] On a minute-by-minute basis AC units are either on or off even on hot days. As the temperature within a home rises above the thermostat setting, the AC unit turns on and transfers heat from inside a home to the outside. In practice, however, several days and weeks can pass by without a heat wave. As a result, any uncapped in-day adjustment using individual AC data can be highly volatile because the adjustment is based on the ratio of actual use during pre-event days and baseline. If the AC was off for several days this leads to a small value in the denominator – dividing by a very small number produces a large value and large adjustments. The reverse is also true, if an AC unit was not in operation during pre-event hours, the actual use value is low, leading to a large downward adjustment. Capping adjustments doesn't produce more accurate results because AC use on hotter days is often double or triple the use in prior days and weeks. If baseline methods with in-day adjustments are to be employed with AC end-use data, it is necessary to aggregate across the AC units.

In comparison, all the regression models tested produced highly accurate baseline and impact estimates for the average event. The estimated impacts for the average event range from 0.19 to 0.21 kW per SmartAC household and match closely to the actual impact of 0.20 kW. Differences between the regression model results are very similar, but the most accurate method

[26] Appendix D provides a detailed example of how in-day adjustments are applied.

was the simplest and relies exclusively on external factors such as weather to produce the estimates. It models AC use and impacts as a function of day characteristics, hour and temperatures and does not predict AC use as a function of AC use in the same day or the prior day. As a result, it can also be used to forecast the AC load reduction capability under different weather conditions.

While lack of bias in the measurements is critical, it is not the only criteria for accuracy. It is possible for a settlement alternative method to be inaccurate for individual curtailment hours but is accurate on average because the errors cancel each other out.

Figure 4-1 compares the actual impacts per household impacts by event day with the predictions from the two most accurate settlement alternatives using AC end-use data, the first and second regression models. Both models perform relatively well across most curtailment events, with a few exceptions.

Figure 4-1: Comparison of Actual and Predicted Values by Date
Individual AC End-use Data

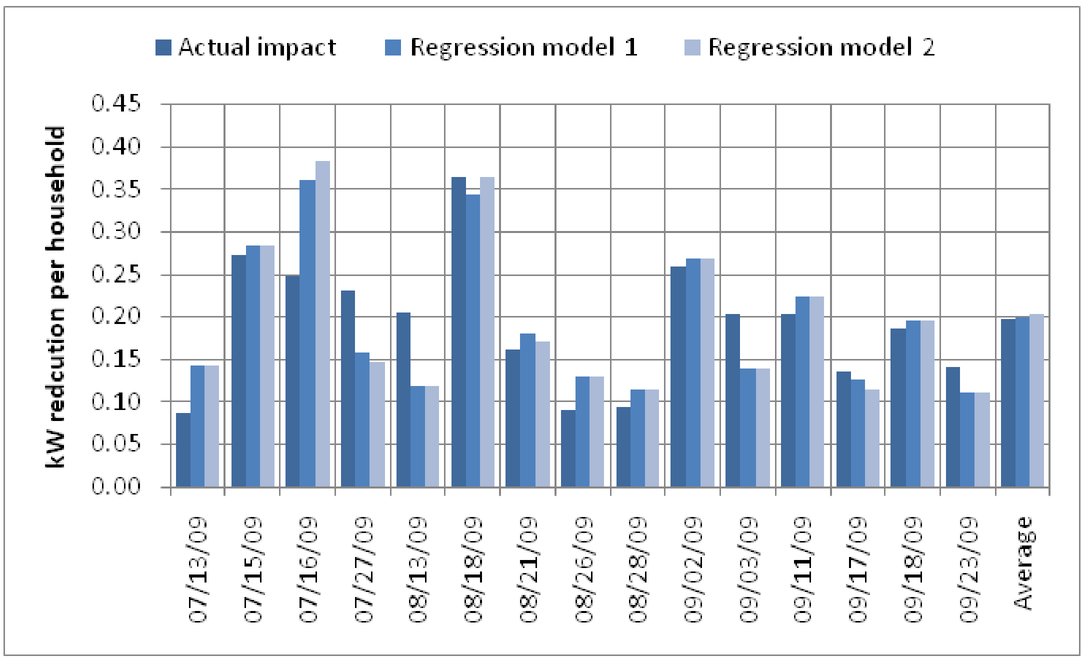


Table 4-3 compares the actual and estimated demand reductions for each curtailment event and each of the within-subject measurement alternatives. It also summarizes several goodness-of-fit statistics that describes how close the measurements were to actual values. For all of the goodness-of-fit metrics in the table, a lower value indicates a higher degree of accuracy. The most intuitive metric is the mean absolute percentage error, which describes the average magnitude of the errors, but does not allow negative and positive errors to cancel each other out. While the most accurate method does not over or under predict overall curtailment hours, it produces a 21% average percent error for individual curtailment hours.

Table 4-3: Accuracy and Goodness-of-Fit of Settlement Alternatives for Each Simulated Event Individual AC End-use Data

Date	Proxy Event Date	Daily Max Temp (°F)	Actual Impacts	Predicted Impacts							
				Day Matching			Weather Matching	Regression			
				1	2	3		1	2	3	4
07/13/09	90.0	94.3	0.09	0.06	0.11	0.17	0.01	0.14	0.14	0.16	0.11
07/15/09	96.4	97.0	0.27	-0.34	0.77	0.51	0.51	0.28	0.28	0.31	0.31
07/16/09	98.1	98.2	0.25	-0.32	0.14	0.10	0.46	0.36	0.38	0.37	0.35
07/27/09	81.9	97.6	0.23	-0.20	0.24	0.22	0.24	0.16	0.15	0.13	0.05
08/13/09	89.3	91.4	0.21	-0.19	0.37	0.31	0.37	0.12	0.12	0.13	0.14
08/18/09	91.6	92.5	0.36	0.09	0.30	0.30	0.27	0.34	0.36	0.33	0.30
08/21/09	80.1	97.3	0.16	-0.07	0.13	0.20	0.07	0.18	0.17	0.18	0.06
08/26/09	87.1	90.7	0.09	0.17	0.02	0.08	0.13	0.13	0.13	0.15	0.17
08/28/09	91.8	98.1	0.09	-0.03	0.14	0.14	0.04	0.11	0.11	0.13	0.15
09/02/09	96.0	97.3	0.26	-0.13	0.64	0.51	0.16	0.27	0.27	0.29	0.26
09/03/09	76.8	97.2	0.20	0.04	0.18	0.20	0.26	0.14	0.14	0.14	0.18
09/11/09	80.8	96.4	0.20	-0.02	0.31	0.25	0.19	0.22	0.22	0.24	0.19
09/17/09	87.5	90.2	0.14	0.05	1.83	2.56	1.50	0.13	0.12	0.12	0.13
09/18/09	79.4	96.6	0.19	0.00	0.43	0.37	0.23	0.20	0.20	0.21	0.19
09/23/09	88.0	94.0	0.14	-0.21	0.32	0.40	0.35	0.11	0.11	0.11	0.07
Average	**87.7**	**95.3**	**0.20**	**-0.07**	**0.38**	**0.39**	**0.31**	**0.20**	**0.20**	**0.21**	**0.19**
Bias	**Mean percent error**			**-135.9%**	**93.7%**	**95.9%**	**55.4%**	**0.8%**	**2.3%**	**6.3%**	**-3.4%**
Goodness of Fit	**Mean absolute error (MAE)**			**0.30**	**0.23**	**0.22**	**0.15**	**0.04**	**0.04**	**0.05**	**0.05**
	Mean absolute percent error (MAPE)			**141.3%**	**128.8%**	**145.7%**	**99.3%**	**23.3%**	**25.6%**	**30.0%**	**31.0%**
	Root mean squared error (RMSE)			**0.36**	**0.44**	**0.53**	**0.31**	**0.05**	**0.06**	**0.06**	**0.07**
	Normalized RMSE (CVRMSE)			**1.82**	**2.24**	**2.72**	**1.64**	**0.26**	**0.29**	**0.30**	**0.35**
	Coefficient of alienation (CoA)			**19.03**	**28.70**	**42.25**	**17.09**	**0.44**	**0.49**	**0.59**	**0.77**
	Chi-squared			**11.14**	**25.74**	**44.86**	**14.74**	**0.28**	**0.34**	**0.39**	**0.53**

4.2.2 Aggregated Air Conditioner End-use Data

Aggregated AC data does not include non-participants or other end-uses and, as a result, the ratio of the signal – the load reduction – to the background noise is larger. When aggregated, the AC end-use data is much more accurate for day and weather-matching baselines. Zero or very small values are less likely with aggregated AC data because it is less likely that all units are off during the days prior to an event or during pre-event periods. This reduces the volatility of the in-day ratio adjustments and produces more accurate estimates. On other hand, impact estimates from regression models are generally less accurate with aggregated rather than individual AC data, although they are still highly accurate.

Table 4-4 summarizes the degree to which each regression, day-matching and weather-matching settlement alternative over or underestimated the demand reduction over 15 curtailment events. It shows if the baseline and impact estimates are biased. The results are presented on a per household basis.

Table 4-4: Accuracy by Settlement Alternative for Average Event Aggregate Air Conditioner Data

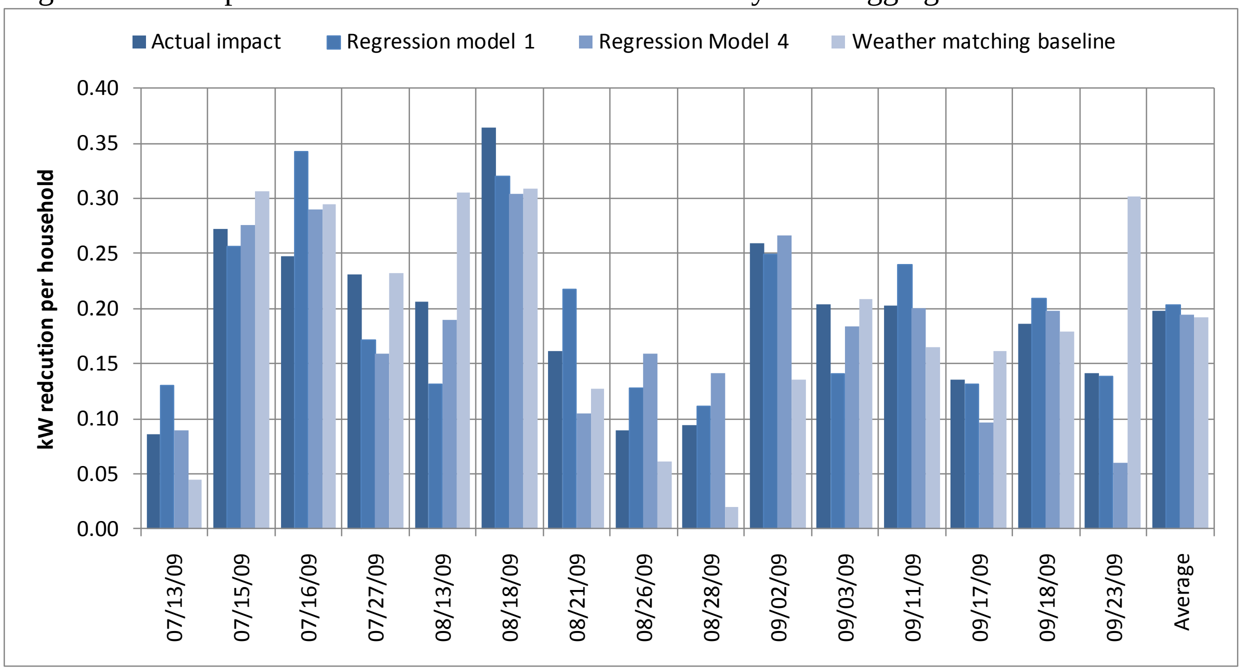
Method	Calculation	Actual load without DR	Predicted load without DR	Mean % Error	Actual Impact	Predicted Impact	Mean % Error
Day Matching	1. 10-in-10 baseline with 20% in-day adjustment cap	0.78	0.57	-26.8%	0.20	-0.01	-105.3%
	2. 10-in-10 baseline with uncapped in-day adjustment	0.78	0.81	3.7%	0.20	0.23	14.4%
	3. Top 3-in-10 baseline with uncapped in-day adjustment	0.78	0.80	2.4%	0.20	0.22	9.2%
Weather Matching	4. Weather baseline with in-day adjustment	0.78	0.77	-0.8%	0.20	0.19	-3.3%
Regression	5. With treatment variables and lags or leads	0.78	0.74	-4.8%	0.20	0.20	2.7%
	6. With treatment variables and a day lag variable	0.78	0.76	-1.6%	0.20	0.21	5.0%
	7. With treatment variables and pre and post event hour variables	0.78	0.79	1.9%	0.20	0.23	14.6%
	8. With no treatment variables and with pre and post event hour variables	0.78	0.77	-0.4%	0.20	0.19	-1.7%

While the underlying data is identical, aggregating the AC end-use data produces marked improvements for day and weather-matching baselines in comparison to relying on individual AC end-use data. The uncapped day and weather-matching baseline estimates produce mean percent errors ranging from -3.3% to 14.4%. This is a substantial improvement over using individual AC data which, by comparison, produced mean percent errors ranging from 55.4% to 95.9% for the same baseline methods. However, the 10-in-10 day-matching baseline with the 20% adjustment cap still performs rather poorly. On average, it reports no demand reductions when in fact the loads dropped by 0.20 kW per household. Simply put, the 20% cap on in-day adjustments is too stringent for residential accounts because AC electricity use is highly weather sensitive.

Despite the improvement in day and weather-matching baselines, as a whole, the 10-in-10 baseline with or without a cap on in-day adjustments still produced the least accurate impacts. The three most accurate settlement alternatives were the weather-matching baselines and Regressions 1 and 4. All of them, on average, produce measurement that, on average, over or underestimate by less than 3.5%. To put this in context, they produced impact estimates between 0.192 kW to 0.203 kW per household for the average event, which compare rather well to the true impacts of 0.198 kW per household.

Figure 4-2 compares the estimates from the three alternatives with the least bias to the actual demand reductions per household by curtailment event.

Figure 4-2: Comparison of Actual and Predicted Values by Date Aggregated AC End-use Data



While the weather-matching baseline does not over or underestimate on average, it leads to substantial errors on specific curtailment events. For example, the impacts for August 13, September 2 and September 23 are 49% too high, 48% too low and 114% too high, respectively. This highlights the need to systematically assess not only the degree of bias but the how well it predicts for individual curtailment periods.

Table 4-5 compares actual and predicted values for all of the within-subject settlement alternatives tested using aggregated AC data. It also shows the metrics for bias and goodness-of-fit introduced in Section 3.6. Regression Model 4 provides the best fit to actual impacts and least amount of error across events. The accuracy metrics are similar across event days for Regression

1. However, the weather-matching baseline is unbiased and is less accurate than the other two options.

Table 4-5: Accuracy and Goodness-of-Fit of Settlement Alternatives for Each Simulated Event Aggregate AC End-use Data

Date	Daily Max Temp (°F)	Actual Impacts	Predicted Impacts							
			Day Matching			Weather Matching	Regression			
			1	2	3		1	2	3	4
07/13/09	94.3	0.09	0.09	0.02	0.00	0.04	0.13	0.14	0.14	0.09
07/15/09	97.0	0.27	-0.14	0.52	0.46	0.31	0.26	0.26	0.31	0.28
07/16/09	98.2	0.25	-0.04	0.33	0.26	0.30	0.34	0.39	0.38	0.29
07/27/09	97.6	0.23	-0.06	0.26	0.31	0.23	0.17	0.18	0.21	0.16
08/13/09	91.4	0.21	-0.12	0.28	0.29	0.31	0.13	0.14	0.15	0.19
08/18/09	92.5	0.36	0.19	0.31	0.25	0.31	0.32	0.35	0.34	0.30
08/21/09	97.3	0.16	-0.04	0.16	0.16	0.13	0.22	0.20	0.24	0.11
08/26/09	90.7	0.09	0.16	0.05	0.04	0.06	0.13	0.13	0.15	0.16
08/28/09	98.1	0.09	0.00	0.06	0.08	0.02	0.11	0.11	0.11	0.14
09/02/09	97.3	0.26	-0.03	0.33	0.31	0.14	0.25	0.25	0.30	0.27
09/03/09	97.2	0.20	0.05	0.20	0.21	0.21	0.14	0.14	0.15	0.18
09/11/09	96.4	0.20	-0.02	0.16	0.16	0.17	0.24	0.23	0.26	0.20
09/17/09	90.2	0.14	0.01	0.12	0.16	0.16	0.13	0.13	0.13	0.10
09/18/09	96.6	0.19	-0.02	0.17	0.17	0.18	0.21	0.21	0.22	0.20
09/23/09	94.0	0.14	-0.24	0.22	0.23	0.30	0.14	0.13	0.14	0.06
Average	**95.5**	**0.20**	**-0.01**	**0.23**	**0.22**	**0.19**	**0.20**	**0.21**	**0.23**	**0.19**
Bias	**Mean percent error (MPE)**		**-105.3%**	**14.4%**	**9.2%**	**-3.3%**	**2.7%**	**5.0%**	**14.6%**	**-1.7%**
Goodness of Fit	**Mean absolute error (MAE)**		**0.22**	**0.06**	**0.06**	**0.05**	**0.04**	**0.04**	**0.05**	**0.04**
	Mean absolute percent error (MAPE)		**111.3%**	**31.8%**	**30.6%**	**28.5%**	**23.7%**	**25.2%**	**29.5%**	**23.6%**
	Root mean squared error (RMSE)		**0.25**	**0.10**	**0.09**	**0.07**	**0.05**	**0.06**	**0.06**	**0.04**
	Normalized RMSE (CVRMSE)		**1.26**	**0.50**	**0.43**	**0.35**	**0.26**	**0.29**	**0.31**	**0.22**
	Coefficient of alienation (CoA)		**10.79**	**1.70**	**1.27**	**0.82**	**0.46**	**0.58**	**0.66**	**0.33**
	Chi-squared		**6.10**	**0.82**	**0.68**	**0.53**	**0.29**	**0.35**	**0.43**	**0.28**

4.2.3 Household Data

In comparison to the AC end-use data, household data includes several other end-uses and has more background noise from which the signal – the AC control load reductions – must be detected. In hotter days, the amount of AC load is a high proportion of overall household electricity demand. In cooler days, it is a small share of the overall household electricity demand, making it more difficult to detect AC control impacts from the background noise.

For the analysis, the summer 2010 data was sampled for participants in 204 randomly selected feeders.[27] Within each feeder, if more than 100 households on the feeder were enrolled in SmartAC, smart meter data from a randomly selected group of 100 participating households was employed. If there were less than 100 participants in a feeder, all SmartAC household with smart meter data were sampled. Because at the time of analysis, the PG&E smart meter deployment was not fully complete, a total of 132 feeders representing 88.5% of the SmartAC population were employed. Each of these SmartAC households had smart meter data for the entire 2010 summer (May to October) and their data was aggregated by feeder. This approach was selected to enable comparison for the same feeders between smart meter and feeder data.

Table 4-6: Accuracy by Settlement Alternative for Average Event Household Data

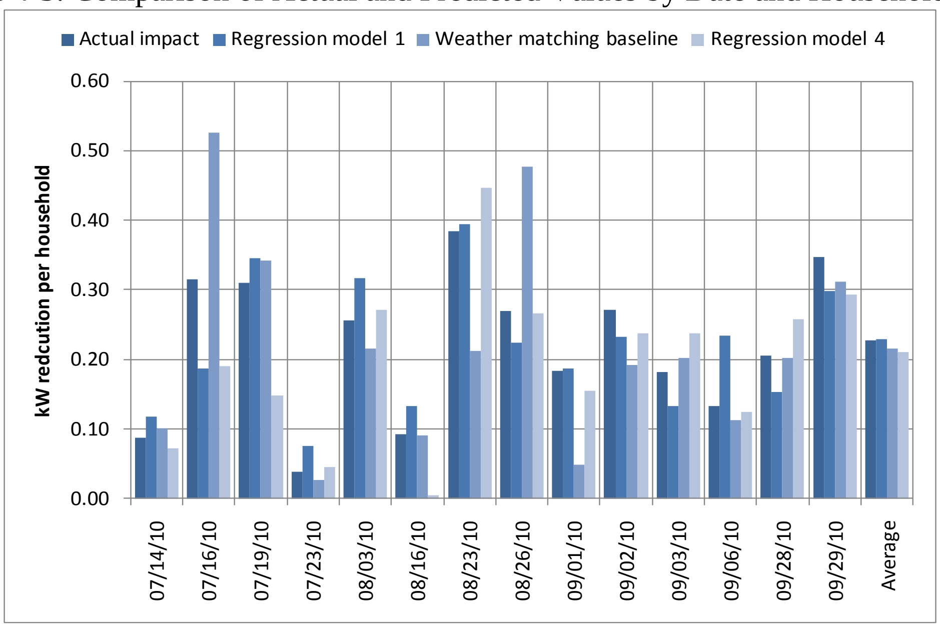
Method	Calculation	Actual load without DR	Predicted load without DR	Mean % Error	Actual Impact	Predicted Impact	Mean % Error
Day Matching	1. 10-in-10 baseline with 20% in-day adjustment cap	2.22	1.94	-12.6%	0.23	-0.05	-122.7%
	2. 10-in-10 baseline with uncapped in-day adjustment	2.22	2.14	-3.7%	0.23	0.15	-35.9%
	3. Top 3-in-10 baseline with uncapped in-day adjustment	2.22	2.19	-1.1%	0.23	0.20	-10.8%
Weather Matching	4. Weather baseline with in-day adjustment	2.22	2.20	-0.6%	0.23	0.21	-5.8%
Regression	5. With treatment variables and lags or leads	2.22	2.14	-3.6%	0.23	0.23	0.2%
	6. With treatment variables and a day lag variable	2.22	2.20	-0.7%	0.23	0.25	11.6%
	7. With treatment variables and pre and post event hour variables	2.22	2.22	0.3%	0.23	0.26	13.5%
	8. With no treatment variables and with pre and post event hour variables	2.22	2.20	-0.8%	0.23	0.21	-7.7%

[27] The share of customers with smart meters in 2009 was far smaller and, as a result, the analysis of AC end-use data is based on 2009 data while the analysis of household and feeder data relies on 2010 data.

Table 4-6 summarizes the degree to which each regression, day-matching and weather-matching settlement alternative over or under predicts the true demand reductions on average. It reflects whether the measurement option is biased. As in prior tables, the results are presented on a per-household basis. On average, Regression Model 1 and weather-adjusted baselines produce results that least over or under predict demand reductions. The 10-in-10 day-matching baseline method with the 20% in-day adjustment cap is the least accurate. Removing the cap improves its accuracy, but does not do so enough. The other notable difference is that the values for the baseline are substantially larger on a per household basis than those using AC end-use data, though demand reductions are similar. This is because the baseline now includes all household end-uses in addition to AC, increasing the background noise.

Figure 4-3 compares the actual impacts per household by event day with the estimates from the three most accurate settlement alternatives using household data – Regression Model 1, the weather-matching baseline and Regression Model 4.

Figure 4-3: Comparison of Actual and Predicted Values by Date and Household Data



The weather-matching baseline is once again accurate for the average event, but produces larger errors on specific event dates than other alternatives. In specific, the impacts for July 16, August 23 and August 26 are 66.9% too high, 44.9% too low and 77.4% too high, respectively. Each of those days had relatively large per household impacts.

Table 4-7 compares actual to predicted values for all the settlement alternatives tested and for each curtailment event. It shows the metrics for assessing goodness-of-fit and bias introduced in Section 3.6.

Table 4-7: Accuracy and Goodness-of-Fit
Across Event Days Within-subject Alternatives Household Data

Date	Daily Max Temp (°F)	Actual Impacts	Predicted Impacts							
			Day Matching			Weather Matching	Regression			
			1	2	3		1	2	3	4
07/14/10	94.3	0.09	-0.02	0.01	0.13	0.10	0.12	0.13	0.15	0.07
07/16/10	97.0	0.31	-0.34	0.56	0.48	0.52	0.19	0.21	0.17	0.19
07/19/10	98.2	0.31	-0.14	0.38	0.40	0.34	0.35	0.37	0.37	0.15
07/23/10	97.6	0.04	0.02	0.03	0.10	0.03	0.07	0.08	0.10	0.04
08/03/10	91.4	0.26	0.22	0.26	0.25	0.22	0.32	0.34	0.33	0.27
08/16/10	92.5	0.09	0.01	0.11	0.05	0.09	0.13	0.15	0.12	0.01
08/23/10	97.3	0.38	0.01	0.19	0.27	0.21	0.39	0.42	0.43	0.45
08/26/10	90.7	0.27	0.39	0.45	0.46	0.48	0.22	0.25	0.24	0.27
09/01/10	98.1	0.18	0.11	0.25	0.17	0.05	0.19	0.23	0.19	0.16
09/02/10	97.3	0.27	-0.10	-0.01	0.17	0.19	0.23	0.25	0.27	0.24
09/03/10	97.2	0.18	0.28	0.31	0.20	0.20	0.13	0.19	0.23	0.24
09/06/10	96.4	0.13	0.04	0.06	0.20	0.11	0.24	0.26	0.26	0.13
09/28/10	90.2	0.21	-0.27	-0.18	-0.08	0.19	0.15	0.17	0.20	0.26
09/29/10	96.6	0.35	-0.71	-0.03	0.17	0.31	0.30	0.33	0.33	0.29
Average	**95.3**	**0.23**	**-0.05**	**0.15**	**0.20**	**0.22**	**0.23**	**0.26**	**0.26**	**0.21**
Bias	**Mean percent error (MPE)**		**-122.8%**	**-36.0%**	**-10.6%**	**-5.7%**	**0.4%**	**11.8%**	**13.6%**	**-7.5%**
Goodness of Fit	**Mean absolute error (MAE)**		**0.31**	**0.16**	**0.10**	**0.06**	**0.05**	**0.05**	**0.05**	**0.05**
	Mean absolute percent error (MAPE)		**120.2%**	**69.5%**	**48.6%**	**25.3%**	**29.3%**	**31.7%**	**35.8%**	**24.4%**
	Root mean squared error (RMSE)		**0.43**	**0.21**	**0.13**	**0.09**	**0.06**	**0.06**	**0.06**	**0.07**
	Normalized RMSE (CVRMSE)		**1.88**	**0.92**	**0.57**	**0.40**	**0.25**	**0.27**	**0.28**	**0.30**
	Coefficient of alienation (CoA)		**17.18**	**4.15**	**1.60**	**0.79**	**0.30**	**0.34**	**0.39**	**0.45**
	Chi-squared		**13.71**	**3.84**	**1.67**	**0.67**	**0.43**	**0.55**	**0.64**	**0.45**

As in all prior analysis, the 10-in-10 day-matching baseline with the 20% cap on in-day adjustments performs the worst. While it does not over or under predict on average, based on the goodness-of-fit metrics, the weather-matching baseline is once again less accurate than regression alternatives.

Overall, the best settlement alternatives with household data are as good if not better than the best alternatives with AC end-use data. This can be seen by comparing the normalized goodness-of-fit and bias metrics in Tables 4-3, 4-5 and 4-7. The mean percent error for the best option using household data, 0.4%, is comparable to best options with AC individual and aggregated data, 0.8% and -1.7%, respectively. The goodness-of-fit metrics for the best option with household data are less favorable in some cases than the comparable AC data metrics – e.g., MAPE and CVRMSE – and better for other metrics such as the coefficient of alienations (which measures the amount of unexplained variation). This is important. It leads to the conclusion that estimating impacts with smart meter data is not only less costly, but nearly as good as analyzing directly metered end-use data.

The use of household data from smart meters as a basis for settlement analyses comes with several distinct advantages. In California and many areas of the U.S., smart meters will be or already are universally available for residential and commercial customers. It also provides the ability to use large samples at relatively low cost. If the AC control-devices can be individually addressed, as is the case for PG&E and for most newer AC load control systems, it is possible to randomly assign a sub-set of the population to act as a control group.

4.2.4 Feeder Data

Using feeder data has garnered a substantial amount of attention primarily because the data collection system is already in place, can provide near real time visibility and provides the ability to extract data remotely. The demonstrations testing the ability to use AC load control for operations relied on feeders that had among the highest share of households enrolled in load control programs. However, they were specially selected feeders.

A systematic analysis of randomly selected feeders better reflects whether or not feeder data can be used for settlement or operations. Overall, a random sample of 10% of the feeders with controllable AC units was requested for analysis. However, of the 204 feeders sampled, only 85 had hourly or sub-hourly data that could be used for measuring demand reductions. In addition, as described in Appendix B, those feeders were skewed to the Greater Bay Area and not all of them had data available for the full 2010 summer. In other words, hourly or sub-hourly feeder data is not always available for the full population and there are often systematic differences between areas that do and do not have feeder data available. This limits the ability to use feeder data to measure demand reductions for operations and settlement.

Table 4-8 summarizes the characteristics of the sampled feeders for each 2010 simulated curtailment event. The columns on the left side describe the characteristics of the average feeder, including the overall feeder loads, the controllable AC load and the aggregate curtailment from those AC loads. The columns on the right side show the distribution of the demand reduction as a percent of feeder loads across the 85 feeders with data.

Table 4-8: Distribution of Percent Impacts on Feeders

Date	Average Feeder					Load per AC unit	Distribution of % Impacts on Feeder (50% AC Cycling)			
	Load without DR	AC load of SmartAC participants	Actual Aggregate Impact	% Impact	Temp		50th	75th	95th	Max
	(kW)	(kW)	(kW)	(%)	(°F)	(kW)	(%)	(%)	(%)	(%)
7/14/2010	6,211.9	18.2	4.2	0.07%	82.5	0.19	0.01%	0.10%	0.35%	0.95%
7/16/2010	6,840.5	57.3	19.2	0.28%	74.9	0.50	0.06%	0.31%	1.25%	1.49%
7/19/2010	7,392.2	58.0	16.4	0.22%	82.4	0.57	0.02%	0.26%	0.98%	1.66%
7/23/2010	5,796.6	13.1	1.8	0.03%	77.9	0.14	0.00%	0.04%	0.16%	0.56%
8/3/2010	6,548.4	41.5	14.0	0.21%	80.4	0.41	0.04%	0.28%	1.03%	1.45%
8/16/2010	6,122.4	25.0	4.8	0.08%	70.2	0.26	0.01%	0.12%	0.38%	0.45%
8/23/2010	7,628.3	82.1	27.2	0.36%	94.3	0.79	0.16%	0.38%	1.48%	1.95%
8/26/2010	6,503.0	41.8	13.8	0.21%	73.3	0.44	0.05%	0.34%	0.91%	1.19%
9/1/2010	7,128.1	50.8	13.6	0.19%	80.7	0.47	0.08%	0.23%	0.71%	0.95%
9/2/2010	7,879.9	55.5	17.7	0.22%	94.6	0.51	0.08%	0.22%	1.05%	1.65%
9/3/2010	5,687.3	32.2	11.6	0.20%	69.2	0.27	0.05%	0.27%	0.89%	1.41%
9/6/2010	6,313.9	40.5	8.8	0.14%	90.9	0.36	0.05%	0.17%	0.56%	0.73%
Total	**6,671.0**	**43.0**	**12.8**	**0.19%**	**80.9**	**0.41**	**0.04%**	**0.20%**	**0.88%**	**1.95%**

For the average feeder, the curtailment events led to an average reduction of 0.2% for the feeder loads. Even for the feeders with the highest penetration of load control devices, the curtailments rarely exceed more than 1% or 2% of the feeder loads.

For the majority of feeders and days, the AC impacts are relatively small compared to the feeder loads because only a small fraction of the feeder load is controllable. The average feeder in the dataset had roughly 2,200 residential accounts, of which, on average, 85 were enrolled in the AC load control program. In addition, feeders typically include commercial and industrial accounts that may not be eligible for load control and additional non-AC end uses. Feeder data can detect AC load reductions impact for feeders with extremely high saturation levels on very hot days. For most feeders, however, it is difficult to distinguish the load reduction – the "signal" – from normal fluctuations in feeder loads. This is true even when 99% of the variation in feeder loads can be explained. To put this in perspective, the remaining 1% of unexplained variation is five times larger than the demand reduction, 0.2%.

Appendix I includes a more detailed discussion on how the data source affect demand reduction measurement accuracy and includes side-by-side comparisons of feeder loads with more granular data sources.

Table 4-9 compares the actual and estimated demand reductions with each of the calculation methods used for feeder data. It also includes the metrics for assessing bias and goodness-of-fit introduced in Section 3.6. In total, the average event day impact was 0.16 per SmartAC household across sampled feeders with data.[28]

Many of the calculation methods predicted the feeders load very well. Several alternatives produced estimates that over or under estimated feeder loads by less than 1% and explained over 98% of the variation in feeder loads. While baseline and regression accurately predict feeder loads absent DR, they cannot accurately distinguish the demand reductions from normal, but unexplained fluctuations in feeder loads.

Alternatives produce very small baseline errors such as regression model 4 still lead to large errors in the impact estimates.[29] The most accurate results are from the two regression models that make use of same day or prior day load and employ treatment variables. While they are better able to distinguish impacts from baseline error, they are still relatively inaccurate in comparison to alternatives based on AC end use of whole household data and do not perform well for individual event days as can be seen in the goodness-of-fit metrics.

To pressure test the ability of feeder data to produce accurate impacts, the impacts were simulated under the assumption that AC compressors could indeed be fully shut down for short interval given the current PG&E customer agreement. Doing so more than doubles the impact per household and produces a stronger signal, making it easier to detect AC load reductions from

---

[28] The impacts per household are lower because less feeder data was not available in the hotter regions. In addition, the SCADA was not available for all feeders through the last week of September. As a result, feeder data was available for 12 of the simulated events.

[29] Regression model 4 had a baseline MPE of-0.1% and explained 98.3% of the variation in feeder loads.

the background noise. Table 4-10 presents the accuracy of the different baseline calculation methods with feeder data and larger AC load reductions.

Table 4-9: Accuracy and Goodness-of-Fit of Within Subject Settlement Alternatives
Feeder Data with 50% AC Cycling

Date	Daily Max Temp (°F)	Actual Impacts per home	Predicted Impacts							
			Day Matching			Weather Matching	Regression			
			1	2	3		1	2	3	4
07/14/10	94.3	0.05	-2.07	-1.54	-0.68	0.74	-0.10	-0.28	-0.12	0.23
07/16/10	97.0	0.23	-2.54	4.88	3.64	3.42	-0.18	-0.13	-0.27	2.50
07/19/10	98.2	0.20	-3.18	0.91	1.64	0.41	-0.28	-0.41	-0.34	-1.44
07/23/10	97.6	0.02	-0.43	-0.41	0.49	-0.55	-0.15	-0.24	-0.08	-0.05
08/03/10	91.4	0.17	-0.61	0.07	-0.29	0.20	0.79	0.87	0.73	0.54
08/16/10	92.5	0.06	0.38	0.48	-0.53	0.00	-0.10	-0.06	-0.15	0.02
08/23/10	97.3	0.33	-9.35	-4.25	-3.78	-4.90	-0.40	-0.62	-0.45	-0.54
08/26/10	90.7	0.17	1.33	1.91	1.33	2.75	0.32	0.32	0.25	1.01
09/01/10	98.1	0.16	-2.58	2.07	0.23	-3.22	-1.14	-1.22	-1.63	-2.38
09/02/10	97.3	0.21	-9.26	-6.55	-2.64	-2.50	0.25	-0.11	0.00	-0.54
09/03/10	97.2	0.14	2.03	2.69	0.69	1.28	0.02	0.28	0.67	1.31
09/06/10	96.4	0.11	-1.74	-2.29	0.91	-2.04	0.51	0.31	0.48	0.33
Average	**95.7**	**0.16**	**-2.75**	**-0.68**	**-0.17**	**-0.59**	**0.02**	**-0.07**	**0.00**	**0.05**
Bias	**Baseline percent error**		**-3.6%**	**-1.0%**	**-0.4%**	**-0.9%**	**-9.4%**	**-1.4%**	**-0.2%**	**-0.1%**
	Mean percent error (MPE)		**-1810%**	**-526%**	**-203%**	**-463%**	**-89%**	**-140%**	**-98%**	**-72%**
Goodness of Fit	**Mean absolute error (MAE)**		**3.46**	**2.46**	**1.45**	**1.85**	**0.40**	**0.46**	**0.48**	**0.88**
	Mean absolute percent error (MAPE)		**21.14**	**16.63**	**9.32**	**11.65**	**3.04**	**3.67**	**3.31**	**5.49**
	Root mean squared error (RMSE)		**4.80**	**3.23**	**1.97**	**2.48**	**0.51**	**0.58**	**0.64**	**1.17**
	Normalized RMSE (CVRMSE)		**29.84**	**20.10**	**12.27**	**15.30**	**3.17**	**3.62**	**4.01**	**7.26**
	Coefficient of alienation (CoA)		**3,125.17**	**1,417.72**	**528.68**	**830.05**	**35.23**	**46.03**	**56.47**	**184.97**
	Chi-squared		**1,937.47**	**1,014.99**	**323.27**	**535.34**	**30.24**	**41.03**	**44.42**	**139.85**

Table 4-10: Accuracy and Goodness-of-Fit of Within Subject Settlement Alternatives Feeder Data with 100% AC Cycling

Date	Daily Max Temp (°F)	Actual Impacts per home	Predicted Impacts							
			Day Matching			Weather Matching	Regression			
			1	2	3		1	2	3	4
07/14/10	94.3	0.17	6.12	-1.07	0.02	0.82	0.02	0.07	0.08	1.18
07/16/10	97.0	0.54	3.47	5.31	4.18	3.36	0.05	0.29	-0.01	2.76
07/19/10	98.2	0.55	8.64	0.63	1.36	0.04	0.10	0.57	-0.03	-0.53
07/23/10	97.6	0.12	1.00	-0.26	0.43	-0.98	-0.38	-0.72	-0.06	1.15
08/03/10	91.4	0.40	0.59	0.55	0.21	1.25	0.87	1.28	1.13	1.54
08/16/10	92.5	0.23	-20.79	-22.90	-0.43	2.76	0.05	0.29	0.08	1.73
08/23/10	97.3	0.80	-3.85	-4.39	-3.92	-5.52	0.44	1.39	1.06	1.59
08/26/10	90.7	0.39	4.82	2.90	2.19	4.98	0.47	0.57	0.52	1.90
09/01/10	98.1	0.49	-21.15	2.69	-3.15	-3.13	-0.50	0.74	-0.40	-2.25
09/02/10	97.3	0.53	-21.99	-5.59	-1.70	9.59	0.37	0.60	0.44	0.54
09/03/10	97.2	0.31	-16.78	-6.66	-8.34	1.25	0.21	0.10	0.83	2.76
09/06/10	96.4	0.39	-13.07	-2.52	0.63	-10.93	0.57	0.50	0.68	1.29
Average	**95.7**	**0.42**	**-5.96**	**-2.65**	**-1.07**	**0.00**	**0.26**	**0.54**	**0.45**	**1.16**
Bias	**Baseline percent error**		**-7.3%**	**-3.1%**	**-1.4%**	**-0.2%**	**-14.6%**	**-3.2%**	**1.0%**	**0.9%**
	Mean percent error (MPE)		**-1496%**	**-737%**	**-356%**	**-100%**	**-40%**	**25%**	**5%**	**171%**
Goodness of Fit	**Mean absolute error (MAE)**		**10.43**	**4.22**	**2.39**	**3.91**	**0.34**	**0.34**	**0.39**	**1.27**
	Mean absolute percent error (MAPE)		**2869%**	**1310%**	**636%**	**928%**	**102%**	**118%**	**111%**	**373%**
	Root mean squared error (RMSE)		**13.09**	**6.66**	**3.74**	**5.45**	**0.43**	**0.47**	**0.46**	**1.50**
	Normalized RMSE (CVRMSE)		**30.64**	**16.01**	**8.99**	**12.86**	**1.00**	**1.11**	**1.09**	**3.51**
	Coefficient of alienation (CoA)		**4,629.50**	**1,252.95**	**395.62**	**830.95**	**4.97**	**6.04**	**5.84**	**60.76**
	Chi-squared		**8,886.17**	**3,153.41**	**837.45**	**1,265.14**	**9.74**	**15.16**	**10.61**	**119.62**

While the measurements improve, the errors for individual curtailments are, on average, larger than the curtailments. In all cases, settlement alternatives that rely on household or AC end-use data outperform the measurements with feeder data by orders of magnitude. All of the settlement alternatives with feeder data produce large differences between actual and estimated impacts. Simply put, using feeder data for settlement of AC load control programs is highly inaccurate and impractical. Except for highly selected feeders, the percent load impacts on feeder data are extremely small, making it difficult to accurately distinguish reductions in electricity use due to load control from variation in other feeder loads.

4.3 Accuracy with Control Groups

This section presents the results of calculation methods that rely on control groups. In the simulation, customers were randomly assigned to a curtailment or a control group to ensure that the only systematic difference between the two groups is whether or not they were curtailed.

To date, discussions on settlement of DR in electricity markets have focused almost exclusively on day-matching baselines and other algorithms to establish what electricity use would have been in the absence of the demand reduction. This is an outgrowth of the fact that demand response programs for large C&I customers were the first type of DR integrated into electricity markets. It is difficult to employ random assignment with larger customers due to their limited number and the large amount of variation between them. However, with AC direct control, there are a far larger number of participants and less variation in their scale – at least in comparison to large C&I customers. As result, it is possible to use random assignment to estimate AC load impacts using either AC end-use or household data.

This section separately presents the result of using control groups with AC end-use and household data. In both cases, the results are presented using samples with 100, 200, 300, 500, 1,000 and 2,000 sample points each who do and do not have their AC units controlled. In addition, the results are presented using on simple comparison of means and a weather matched difference-in-differences calculation. With the first approach, demand reductions are estimated as the difference between the group that did not have their AC loads curtailed and one that did. With the second approach, the difference also between the two groups is calculated for the curtailment day. However, in addition, differences between the two groups observed during days without curtailments and similar weather are subtracted out. This added step nets out differences that are irrelevant and mainly due to sampling variation and improves the precision of the measurement, particularly if smaller samples are employed. Section 3.5.4 provides more detail about these calculation methods, including examples.

4.3.1 Air Conditioner End-use Data

Table 4-11 summarizes the results using a simple comparison of means for different sample sizes. In each case, we show the distribution of per household load without load control, per household known impacts, and the metrics introduced in Section 3.6 to assess bias and goodness-of-fit. In each case, the table presents the distribution of these metrics across all 100 simulations. In practice, a sample is typically only selected once per year. The tables indicate that likelihood that the randomly assigned samples will yield different levels of accuracy based on the 100 random samples drawn for each sample size category.

The use of control groups and random assignment produces results that do not over or under estimate demand reductions. This is reflected by the mean percent error, a metric for bias, which ranges from -4.1 % to 0.9% for the average sample, depending on sample size.

Table 4-11: Accuracy and Goodness-of-Fit Metrics Random Assignment with a Simple Comparison of Means AC End-use Data with 50% AC Cycling

Sample Size	Metric		Average	Std. Deviation	Distribution Percentiles				
					5th	25th	Median	75th	95th
100	Impacts	Sample True Load w/o DR	0.82	0.07	0.69	0.77	0.81	0.86	0.91
		Sample True Impact	0.21	0.02	0.19	0.20	0.21	0.23	0.24
	Bias	Baseline percent error	-1.1%	11.2%	-17.1%	-7.7%	-1.8%	5.8%	15.1%
		MPE (Demand Reduction)	-4.1%	43.5%	-66.2%	-30.2%	-7.2%	22.6%	58.9%
	Goodness of Fit	MAE	0.13	0.04	0.08	0.10	0.12	0.14	0.19
		MAPE	68.6%	24.6%	43.1%	54.4%	64.8%	75.3%	109.2%
		CVRMSE	0.75	0.24	0.48	0.59	0.69	0.86	1.16
		Coefficient of alienation	3.65	2.77	1.32	2.00	2.89	4.67	7.57
300	Impacts	Sample True Load w/o DR	0.83	0.04	0.76	0.81	0.83	0.86	0.90
		Sample True Impact	0.21	0.01	0.20	0.21	0.21	0.22	0.24
	Bias	Baseline percent error	0.2%	7.6%	-12.1%	-4.5%	-0.4%	4.7%	13.0%
		MPE (Demand Reduction)	0.7%	29.7%	-46.9%	-17.2%	-1.4%	18.3%	51.3%
	Goodness of Fit	MAE	0.08	0.03	0.05	0.06	0.07	0.09	0.14
		MAPE	41.3%	12.7%	26.6%	32.0%	38.8%	48.1%	66.5%
		CVRMSE	0.48	0.15	0.29	0.36	0.44	0.57	0.76
		Coefficient of alienation	1.57	1.00	0.50	0.85	1.23	2.04	3.66
500	Impacts	Sample True Load w/o DR	0.83	0.03	0.78	0.81	0.83	0.85	0.87
		Sample True Impact	0.21	0.01	0.20	0.21	0.21	0.22	0.23
	Bias	Baseline percent error	0.0%	5.4%	-8.3%	-3.6%	-0.6%	3.5%	10.1%
		MPE (Demand Reduction)	-0.1%	20.9%	-32.3%	-14.1%	-2.2%	13.7%	39.0%
	Goodness of Fit	MAE	0.06	0.02	0.04	0.05	0.06	0.07	0.09
		MAPE	31.5%	8.8%	20.2%	25.3%	29.9%	35.2%	50.3%
		CVRMSE	0.36	0.10	0.23	0.29	0.34	0.42	0.55
		Coefficient of alienation	0.90	0.53	0.35	0.53	0.75	1.19	1.94
1000	Impacts	Sample True Load w/o DR	0.83	0.02	0.80	0.82	0.83	0.84	0.86
		Sample True Impact	0.21	0.01	0.21	0.21	0.21	0.22	0.22
	Bias	Baseline percent error	0.2%	3.6%	-5.7%	-2.1%	0.2%	2.1%	6.4%
		MPE (Demand Reduction)	0.9%	14.0%	-22.4%	-8.1%	0.7%	8.3%	24.9%
	Goodness of Fit	MAE	0.04	0.01	0.03	0.04	0.04	0.05	0.07
		MAPE	22.7%	6.4%	14.0%	17.9%	21.3%	26.3%	34.6%
		CVRMSE	0.26	0.07	0.17	0.20	0.24	0.30	0.40
		Coefficient of alienation	0.46	0.28	0.18	0.26	0.37	0.58	0.97
2000	Impacts	Sample True Load w/o DR	0.83	0.01	0.81	0.82	0.83	0.83	0.85
		Sample True Impact	0.21	0.00	0.21	0.21	0.21	0.22	0.22
	Bias	Baseline percent error	-0.1%	2.5%	-4.2%	-1.3%	0.0%	1.3%	3.7%
		MPE (Demand Reduction)	-0.3%	9.6%	-16.4%	-5.2%	-0.2%	4.9%	14.5%
	Goodness of Fit	MAE	0.03	0.01	0.02	0.02	0.03	0.03	0.05
		MAPE	14.8%	4.1%	9.5%	12.5%	14.0%	16.4%	22.2%
		CVRMSE	0.17	0.05	0.11	0.14	0.16	0.19	0.27
		Coefficient of alienation	0.21	0.15	0.08	0.12	0.16	0.24	0.46

However, since samples can only be drawn once, variation from sampling can lead to less accurate measurement and introduce bias. It is more practical to focus on the likelihood of drawing a sample that produces inaccurate estimates. The easiest way to assess the expected bias and accuracy of different samples sizes is by comparing the distribution percentiles. In total, 50% of the samples produced results between the 25th and 75th percentile, and 90% of them produced results between the 5th and 95th percentiles.

For example, when a sample with 500 AC units each in the treatment and control groups was drawn, half the control groups, produced impacts where the bias was between -14.3% and 14.0%, as measured by the MPE. That means that there is approximately a 50% chance that a sample of 500 will over or under estimated demand reductions by more than 14%. With a sample size of 500, over 10% of the control groups produced estimates with a bias that exceeded ±32%. As sample sizes increase, the likelihood of systematic errors due to sampling variation decreases. With samples of 2,000 customers, 50% of control groups drawn produced estimates over or underestimated demand reduction by less than 5%. The same is true for the goodness-of-fit statistics. For example, with a sample of 500 customers there is a 50% chance that the average error for individual events (the MAPE metrics) will exceed 30%, the median. With a large sample of 2,000, there is a 50% chance that the average error for individual events will exceed 16%, the median, and a 5% chance that it will exceed 22%, the 95th percentile.

In practice, samples that collect AC end-use data tend to be costly and typically range between 500 to 1,000. Smaller samples are less reliable, especially they are used to produce demand reduction estimates for more specific areas in the grid. As noted earlier, one of the easiest ways to improve the precision of the measurements with control group is to use the weather matched differences-in-differences calculation.

Table 4-12 summarizes the results if the differences-in-differences calculation is applied. It summarizes the results for the 100 separate samples drawn to reflect role of sampling variation. The relatively simple calculation noticeably improves the precision of the measurements. With a sample of 500, nearly half of control groups produces estimates that over or under reported demand reductions by less 6% and over 90% of the control groups produced estimates with a bias of less than 15%. With larger sample size of 2000, the results improve even more. The likelihood that the control group will substantially over or under estimate the demand reduction is relatively small. The biggest improvement, however, is the goodness of fit statistics, which outperform all other calculation methods so far. There is 50% chance that using a control group of 2,000 accounts will lead to individual curtailment event errors that average less the 12% and that the control group will leave less than 13% of the variation in demand reductions unexplained (this is reflected in the coefficient of alienation).

Table 4-12: Accuracy and Goodness-of-Fit Metrics
Random Assignment with ak Difference-in-differences Calculation
AC End-use Data with 50% AC Cycling

Sample Size	Metric		Average	Std. Deviation	Distribution Percentiles				
					5th	25th	Median	75th	95th
100	Impacts	Sample True Load w/o DR	0.83	0.07	0.71	0.78	0.83	0.88	0.94
		Sample True Impact	0.21	0.02	0.19	0.20	0.21	0.23	0.24
	Bias	Baseline percent error	0.0%	5.1%	-8.5%	-3.5%	-0.1%	3.3%	9.8%
		MPE (Demand Reduction)	0.2%	19.9%	-32.6%	-13.8%	-0.3%	13.0%	38.4%
	Goodness of Fit	MAE	0.11	0.02	0.08	0.09	0.10	0.12	0.15
		MAPE	57.7%	14.6%	38.0%	47.7%	54.8%	64.6%	84.5%
		CVRMSE	0.63	0.16	0.44	0.51	0.61	0.71	0.93
		Coefficient of alienation	2.48	1.18	1.14	1.50	2.23	3.18	4.88
300	Impacts	Sample True Load w/o DR	0.83	0.04	0.77	0.80	0.83	0.85	0.90
		Sample True Impact	0.21	0.01	0.20	0.21	0.21	0.22	0.24
	Bias	Baseline percent error	0.2%	2.8%	-4.1%	-1.8%	0.2%	1.9%	4.6%
		MPE (Demand Reduction)	0.6%	11.1%	-16.0%	-7.0%	0.9%	7.4%	18.0%
	Goodness of Fit	MAE	0.06	0.01	0.05	0.06	0.06	0.07	0.09
		MAPE	34.1%	8.3%	23.7%	29.4%	32.3%	37.4%	48.0%
		CVRMSE	0.39	0.10	0.27	0.32	0.37	0.43	0.58
		Coefficient of alienation	1.02	0.53	0.41	0.64	0.87	1.14	2.31
500	Impacts	Sample True Load w/o DR	0.83	0.03	0.78	0.81	0.83	0.85	0.88
		Sample True Impact	0.21	0.01	0.20	0.21	0.21	0.22	0.23
	Bias	Baseline percent error	0.0%	2.3%	-4.1%	-1.3%	0.0%	1.5%	3.5%
		MPE (Demand Reduction)	0.1%	9.0%	-16.1%	-5.1%	0.0%	6.0%	13.6%
	Goodness of Fit	MAE	0.05	0.01	0.04	0.04	0.05	0.06	0.07
		MAPE	26.4%	5.4%	17.9%	21.8%	26.2%	30.1%	35.6%
		CVRMSE	0.30	0.07	0.20	0.25	0.30	0.34	0.43
		Coefficient of alienation	0.62	0.28	0.27	0.42	0.56	0.75	1.21
1000	Impacts	Sample True Load w/o DR	0.83	0.02	0.81	0.82	0.83	0.84	0.86
		Sample True Impact	0.21	0.01	0.21	0.21	0.21	0.22	0.22
	Bias	Baseline percent error	-0.1%	1.8%	-3.2%	-1.4%	-0.1%	1.2%	2.7%
		MPE (Demand Reduction)	-0.4%	6.9%	-12.6%	-5.6%	-0.2%	4.6%	10.5%
	Goodness of Fit	MAE	0.04	0.01	0.02	0.03	0.04	0.04	0.06
		MAPE	19.3%	4.9%	12.3%	16.0%	18.7%	21.6%	28.4%
		CVRMSE	0.22	0.06	0.14	0.18	0.21	0.24	0.35
		Coefficient of alienation	0.34	0.19	0.13	0.21	0.30	0.37	0.76
2000	Impacts	Sample True Load w/o DR	0.83	0.01	0.81	0.82	0.83	0.84	0.84
		Sample True Impact	0.21	0.00	0.21	0.21	0.21	0.22	0.22
	Bias	Baseline percent error	0.0%	1.0%	-1.7%	-0.7%	-0.1%	0.8%	1.9%
		MPE (Demand Reduction)	0.0%	4.0%	-6.5%	-2.6%	-0.6%	3.1%	7.3%
	Goodness of Fit	MAE	0.02	0.01	0.02	0.02	0.02	0.03	0.03
		MAPE	12.5%	2.6%	8.6%	10.8%	12.1%	14.0%	17.0%
		CVRMSE	0.14	0.04	0.10	0.12	0.14	0.16	0.20
		Coefficient of alienation	0.14	0.07	0.06	0.09	0.13	0.17	0.27

4.3.2 Household Data

While random assignment of control operations is generally accurate with AC end-use data, household data includes other end-uses, producing more background noise from which impacts must be detected. The same process of conducting 100 sampling iterations was employed with household data. There are two main differences between AC end-use and household data. First, with smart meters in place, data collections for equivalent sample sizes are substantially lower for household data than for AC end-us data. Or, to put it differently, sample sizes with household data can be magnitudes larger with the same or lower costs. Second, the percent load reductions with household data are smaller due to the additional background noise from other non-cooling related household end-uses. In specific, the average percent impacts ion household data are 11% compared to 25% for AC end-use data.

Table 4-13 summarizes the results with the simple comparison of means for different sample sizes. In each case we show the distribution of per household load without load control, per household known impacts, and the metrics introduces in Section 3.6 to assess bias and goodness-of-fit. The table indicates the likelihood that a sample drawn will yield different levels of bias and fit.

Not surprisingly, in comparison to the AC data, the household data produces less accurate results. This is exclusively due to the additional noise from end-uses unrelated to AC or cooling. Even when there are 2,000 customers the curtailment and control groups, the impacts can be relatively inaccurate due to chance. In fact, half of the sample draws had errors that underestimated impacts by more than 16% or over overestimated impacts by more than 11%. Likewise, there is 50% change that average error for individual events (MAPE) will exceed 27%.

Table 4-14 summarizes the results using the differences-in-differences calculation. The accuracy of the results improves noticeably in contrast to the simple comparison of means calculation. With 2,000 customers in each group, there is a 50% chance that the samples drawn will, on average, over or under estimate demand reductions by less than 4%. As with AC data, the largest improvement is in the accuracy of impacts for individual event days.

The weighted absolute mean percent error for individual days was less than 22.4% in 90% of sample draws and below 15.7% in half of them. Likewise, the R-squared statistics are noticeably better. They exceed 0.70 in 95% of the sample draws and are over 0.84 in half of them. These results are substantially better than any of the within subjects regression, day-matching or weather-matching baselines. In addition, the household data samples can be increased by orders of magnitude and more aggressive air conditioner control studies are likely to be used when using AC for spinning reserves. Either of those options drastically improves the accuracy of the results. The goodness-of-fit statistics indicate that the using control groups with samples of 2,000 households and the difference-in-differences calculation outperforms all other measurement options except for large control groups of AC end-use data.

Table 4-13: Accuracy and Goodness-of-Fit Metrics Random Assignment with Simple Comparison of Means for Household Data with 50% AC Cycling

Sample Size	Metric		Average	Std. Deviation	Distribution Percentiles				
					5th	25th	Median	75th	95th
100	Impacts	Sample True Load w/o DR	2.00	0.16	1.75	1.88	2.01	2.11	2.27
		Sample True Impact	0.23	0.01	0.22	0.23	0.23	0.24	0.25
	Bias	Baseline percent error	-0.4%	11.1%	-19.1%	-9.1%	-0.3%	7.5%	18.8%
		MPE (Demand Reduction)	-8.3%	97.6%	-176.0%	-87.7%	-2.8%	61.5%	150.1%
	Goodness of Fit	MAE	0.25	0.10	0.12	0.17	0.22	0.31	0.45
		MAPE	140.1%	54.8%	66.6%	100.2%	127.4%	172.6%	240.4%
		CVRMSE	1.26	0.45	0.63	0.88	1.19	1.57	2.08
		Coefficient of alienation	8.17	5.74	1.86	3.50	6.43	11.31	20.17
300	Impacts	Sample True Load w/o DR	2.04	0.11	1.90	1.96	2.03	2.11	2.22
		Sample True Impact	0.23	0.01	0.22	0.23	0.23	0.24	0.24
	Bias	Baseline percent error	1.6%	6.9%	-7.4%	-3.1%	1.3%	5.2%	13.1%
		MPE (Demand Reduction)	12.5%	58.9%	-64.7%	-27.1%	11.6%	43.9%	108.6%
	Goodness of Fit	MAE	0.15	0.07	0.08	0.10	0.13	0.16	0.26
		MAPE	82.8%	39.1%	44.4%	61.0%	74.6%	92.6%	147.6%
		CVRMSE	0.75	0.31	0.44	0.54	0.69	0.85	1.29
		Coefficient of alienation	2.98	3.14	0.89	1.32	2.12	3.38	7.61
500	Impacts	Sample True Load w/o DR	2.01	0.08	1.88	1.95	1.99	2.06	2.14
		Sample True Impact	0.23	0.00	0.23	0.23	0.23	0.24	0.24
	Bias	Baseline percent error	-0.3%	5.3%	-8.1%	-4.3%	-0.8%	2.7%	9.7%
		MPE (Demand Reduction)	-3.9%	45.4%	-72.4%	-38.3%	-7.1%	22.0%	78.2%
	Goodness of Fit	MAE	0.11	0.04	0.06	0.08	0.10	0.14	0.21
		MAPE	64.5%	23.9%	35.8%	45.7%	58.2%	79.8%	105.6%
		CVRMSE	0.58	0.20	0.33	0.42	0.54	0.72	0.97
		Coefficient of alienation	1.69	1.18	0.49	0.80	1.32	2.35	4.31
1000	Impacts	Sample True Load w/o DR	2.02	0.06	1.93	1.99	2.01	2.06	2.13
		Sample True Impact	0.23	0.00	0.23	0.23	0.23	0.24	0.24
	Bias	Baseline percent error	0.4%	4.0%	-6.4%	-1.8%	0.5%	2.9%	6.6%
		MPE (Demand Reduction)	2.9%	34.1%	-57.4%	-16.3%	4.1%	24.5%	55.9%
	Goodness of Fit	MAE	0.08	0.04	0.05	0.06	0.07	0.09	0.16
		MAPE	45.6%	20.7%	25.4%	30.9%	39.5%	53.0%	91.7%
		CVRMSE	0.42	0.17	0.24	0.31	0.37	0.50	0.73
		Coefficient of alienation	0.94	0.87	0.27	0.42	0.62	1.15	2.39
2000	Impacts	Sample True Load w/o DR	2.01	0.03	1.95	1.99	2.01	2.03	2.06
		Sample True Impact	0.23	0.00	0.23	0.23	0.23	0.23	0.24
	Bias	Baseline percent error	-0.4%	2.6%	-4.6%	-1.8%	-0.3%	1.3%	3.8%
		MPE (Demand Reduction)	-3.6%	22.4%	-40.8%	-16.0%	-2.4%	11.3%	32.3%
	Goodness of Fit	MAE	0.06	0.03	0.03	0.04	0.05	0.06	0.11
		MAPE	31.3%	14.0%	16.6%	21.7%	26.6%	36.1%	64.1%
		CVRMSE	0.28	0.11	0.17	0.21	0.24	0.32	0.52
		Coefficient of alienation	0.42	0.39	0.13	0.20	0.27	0.47	1.25

Table 4-14: Random Assignment with Difference-in-differences Accuracy and Goodness-of-Fit for Household Data with 50% AC Cycling

Sample Size	Metric		Average	Std. Deviation	Distribution Percentiles				
					5th	25th	Median	75th	95th
100	Impacts	Sample True Load w/o DR	2.02	0.19	1.76	1.87	2.00	2.16	2.34
		Sample True Impact	0.23	0.01	0.22	0.23	0.23	0.24	0.25
	Bias	Baseline percent error	-0.1%	3.6%	-6.6%	-2.5%	0.2%	1.8%	5.7%
		MPE (Demand Reduction)	-1.4%	31.6%	-54.3%	-21.8%	2.0%	15.6%	49.8%
	Goodness of Fit	MAE	0.17	0.04	0.11	0.14	0.16	0.20	0.25
		MAPE	94.8%	25.7%	56.7%	75.0%	91.6%	110.4%	147.3%
		CVRMSE	0.92	0.22	0.63	0.75	0.90	1.04	1.32
		Coefficient of alienation	4.01	1.99	1.81	2.64	3.68	5.08	7.34
300	Impacts	Sample True Load w/o DR	2.01	0.10	1.86	1.95	2.01	2.07	2.19
		Sample True Impact	0.23	0.01	0.22	0.23	0.23	0.24	0.24
	Bias	Baseline percent error	0.0%	2.0%	-3.2%	-1.6%	-0.2%	1.3%	3.8%
		MPE (Demand Reduction)	-0.2%	17.5%	-26.1%	-13.7%	-2.1%	11.0%	34.5%
	Goodness of Fit	MAE	0.10	0.02	0.07	0.09	0.10	0.12	0.15
		MAPE	57.0%	13.5%	37.6%	48.3%	56.1%	62.9%	84.4%
		CVRMSE	0.55	0.12	0.37	0.47	0.55	0.63	0.75
		Coefficient of alienation	1.44	0.65	0.65	0.98	1.35	1.83	2.53
500	Impacts	Sample True Load w/o DR	2.01	0.08	1.88	1.98	2.01	2.06	2.13
		Sample True Impact	0.23	0.00	0.23	0.23	0.23	0.24	0.24
	Bias	Baseline percent error	-0.1%	1.6%	-2.5%	-1.0%	0.0%	0.9%	2.4%
		MPE (Demand Reduction)	-0.6%	14.0%	-21.2%	-8.7%	0.0%	7.7%	22.0%
	Goodness of Fit	MAE	0.08	0.02	0.05	0.06	0.08	0.09	0.10
		MAPE	42.1%	8.5%	28.4%	35.8%	41.7%	46.9%	57.5%
		CVRMSE	0.41	0.09	0.28	0.34	0.41	0.46	0.58
		Coefficient of alienation	0.79	0.33	0.36	0.54	0.76	0.94	1.52
1000	Impacts	Sample True Load w/o DR	2.02	0.05	1.93	1.98	2.02	2.05	2.11
		Sample True Impact	0.23	0.00	0.23	0.23	0.23	0.24	0.24
	Bias	Baseline percent error	0.0%	0.9%	-1.5%	-0.6%	0.0%	0.7%	1.4%
		MPE (Demand Reduction)	0.0%	8.1%	-13.3%	-5.0%	-0.3%	5.8%	12.5%
	Goodness of Fit	MAE	0.05	0.01	0.04	0.05	0.05	0.06	0.07
		MAPE	29.8%	6.1%	20.9%	25.4%	29.5%	34.0%	41.5%
		CVRMSE	0.29	0.06	0.21	0.25	0.29	0.32	0.39
		Coefficient of alienation	0.40	0.17	0.20	0.28	0.38	0.46	0.70
2000	Impacts	Sample True Load w/o DR	2.02	0.03	1.97	2.00	2.02	2.04	2.07
		Sample True Impact	0.23	0.00	0.23	0.23	0.23	0.23	0.24
	Bias	Baseline percent error	0.0%	0.7%	-1.1%	-0.5%	-0.1%	0.4%	1.2%
		MPE (Demand Reduction)	-0.4%	5.9%	-9.4%	-4.6%	-0.8%	3.3%	10.2%
	Goodness of Fit	MAE	0.04	0.01	0.03	0.04	0.04	0.04	0.05
		MAPE	21.2%	3.8%	15.1%	18.6%	20.5%	23.2%	27.9%
		CVRMSE	0.21	0.04	0.15	0.19	0.20	0.23	0.27
		Coefficient of alienation	0.20	0.07	0.10	0.16	0.19	0.24	0.30

4.4 Key Findings

On aggregate, the comparison of measurement alternatives produced several relevant insights. First, the standard settlement approach used by the California ISO, a 10-in-10 day matching baseline with 20% same day adjustment cap, is inadequate for highly weather sensitive resources such as AC load control. The calculation method had the worst performance of all alternatives tested regardless of the data source used. Second, feeder data has very limited application for estimating AC load reductions for settlement and operations. It can perform well under very high temperatures if a highly saturated feeder is selected and a full load shed strategy is employed, but it cannot be applied widely because many feeders lack hourly or sub-hourly data and in most cases the demand reduction is a very small share of feeder loads and difficult to distinguish for normal variation in such loads. Third, relatively large control groups are required to accurately estimate demand reduction using a control group, if a simple comparisons of means is employed. However, applying the relatively simple weather matched differences in differences calculation described in Section 3.5.4 leads to substantial improvements in the accuracy of the results. Fourth, the simplest and least technical approach – using impact estimates tables – produces relatively accurate results, on average, and outperforms many of the more sophisticated settlement alternatives. While it lacks the precision needed for grid operations, it can be used to simplify the settlement process.

5. A Settlement Framework

The analysis of how well different settlement alternatives estimate actual impact for both the average event and individual event days narrows the number of options that are viable. Clearly, settlement using the California ISO standard 10-in-10 baseline with a 20% adjustment cap is inadequate for residential AC curtailments regardless of the data sources employed. So are impact estimates that rely on feeder data. Both perform far worse than the relying on simple impact estimate tables. Both are inaccurate for the average event day and for individual event days. Based on the results, it is difficult not to conclude that impact estimates for settlement in electricity markets must rely on either AC end-use data or whole household data. Several methods with both of these data sources are highly accurate.

In addition to the accuracy, cost is a key criteria in selecting the settlement framework. There are four main types of cost associated with settlement and they differ for AC end-use and household data:

- *Sample design and implementation:* These are costs associated with designing the sample, addressing load control devices, if needed, and developing the operations plan for the sample.
- *Data collection system*: These are incremental costs to existing data collection infrastructure. For AC end-use data this includes recruitment participants, procuring data collection devices, installing them, and setting up the communications. For household data, the incremental costs can be as low as instructing devices to collect data at sub-hourly intervals.
- *Transmittal of data*: Depending on the system design, there are data transmittal costs associated with moving data from the data collection device to a central point and aggregating it.
- *Data management and calculation of impacts*: These are the costs associated with implementing the settlement alternative.
- *Verification costs*: These are cost incurred by system operator in verifying the settlement calculations.

Operations to stabilize the grid are relatively infrequent and, typically, short in duration. In most markets, the bulk of the payments is for availability, meaning that generators and other resources need to be ready to ramp up resources quickly on a moment's notice. The fact that using impact estimates tables provides relatively accurate AC load reduction estimates raises several questions. Is it really necessary to use more complex settlement calculations for each individual AC load control activation? Is the incremental accuracy worth it? Does the system operator need to verify the measurement and calculations for each individual event? Can impact estimate tables be used for settlement and calibrated and verified annually?

This section discussed the role of data collection technology in supporting grid operations and proposes a framework for settlement that relies on tables with pre-calculated load reductions per device or household – a deemed saving approach. In addition, it compares the estimated cost range of the proposed framework to other settlement alternatives with household and AC end use data. Under this framework, impact estimate tables would be used to settle operations throughout the year. On an annual basis, the impact estimate tables would be updated, reassessed for accuracy and verified by the system operator. The costs from such an approach are compared for calculating impacts for settlement after each operation using AC end-use and household data.

5.1 The Role of Data Collection Technology

It is critical to separately consider the data source and the attributes of the technology used to collect data. There are multiple data collection options for AC, household and feeders. The functionality of the data collection – factors such as the ability to extract the data remotely, time interval recorded and the ability to see usage pattern in close to real time – can affect operations. Different uses require different capabilities and more functionality means higher costs. For example, while the ability to see usage patterns in near real time is required for California ISO operations, it is not required or reconciled for settlement purposes. Likewise, the ability to retrieve AC use patterns an hour or two earlier may be useful for calibrating forecasts of available resources, but is not a pre-requisite of data collection. Table 5-1 summarizes a number of data technology alternatives and their key functional features.

Table 5-1: Data Source and Data Collection Technology Options and Functionality

Source	Technology[30]	% of Coverage of Existing Data Systems	Data Collection Time Interval	Ability to Extract Data	Real Time Visibility	Signal to Noise Ratio
Feeder	SCADA	50%	5 minutes or less	Limited	Yes	Very low
Whole House	Smart meter	100%	15 minutes (w/o affecting billing systems)	Day late	No	Medium
	Metering w/ daily upload capability	0%	5 minutes or less	Day late	No	
	Metering w/ real time capability	0%	5 minutes or less	1 minute lag	Yes	
	Metering w/ end-of-summer extraction	0%	5 minutes or less	Very limited	No	
AC Unit	Day late metering	0%	5 minutes or less	Day late	No	High
	Near Real time metering	0%	5 minutes or less	1 minute lag	Yes	
	End-of-summer metering	0%	5 minutes or less	Very limited	No	

Using feeder data has garnered a substantial amount of attention primarily because the data collection system is already in place has some desirable qualities. When it is available, it often times provides near real time visibility, measurement for small time intervals (e.g., minutes or

[30] A key aspect of technology is the accuracy of the data recording device. This is very different than the accuracy of the demand reduction estimate. The CPUC requires electric meters to meet a +/- 2%. For smart meters, the independent *PG&E Advanced Metering Assessment Report* conducted by The Structure Group in behalf of the CPUC in 2010 concluded that 611 of the 611 smart meters tested met accuracy standards. [add citation] In contrast 141 of the 147 (96%) of the electromechanical meters tested met the +/- 2% accuracy standard. Most data collection devices

seconds) and the ability to extract data remotely. There are several downsides, however. First, as shown, most feeders do not have sufficient penetration of load control devices to distinguish the AC load reduction from the background “noise” for non-participants and other participant end-uses. Second, in many utilities a substantial share of feeders either lack sub-hourly data collection or the data itself is difficult to extract. In the case of PG&E, about half of the feeders have five minute data available and half do not. It is not uncommon for utilities to have visibility for 20% or less of the feeders in place.

A growing number of utilities have installed or are in the process of installing smart meters. Their near universal deployment presents another existing data collection option that can be used for measurement of AC demand reductions. As of November 2010, over 82% of the 125,000 SmartAC participants had smart meters in place. As a result, it is far more economical to select large representative samples or even use a census of the full population. With feeder data, controllable AC load can be a small component of the total feeder electricity use. With household level data for program participation, the controllable AC load is a far larger share of the electricity use. Because PG&E AC control devices are capable of two-way communication, with smart meters it is possible to employ randomly assigned control groups – widely considered the gold standard for estimating impacts accurately – for settlement. However, smart meters also have drawbacks. They normally collect data at hourly intervals but often can be remotely adjusted to collect data for smaller time intervals. From a practical standpoint, changing the smart meter time interval of data collection can require altering data collection or billing systems and impose substantial costs. In the case of PG&E, the utility can remotely change the time interval of data collection to 15-minute measurements without affecting those systems. Using a coarser time interval does affect the ratio between AC electricity use reductions, the "signal," and other electricity use, the "background noise." For example, if the true percent reduction in household load was 30% over a 10-minute dispatch period, with 15-minute intervals the impact would show up as roughly 20%. While not optimal, even after factoring in the larger time intervals, the signal - the percent load reduction - is larger with household level data than it is with feeder data.

While there are other alternatives beside using existing smart meters or feeder data collection system, they rely more so on samples and require deploying new data collection systems and equipment and can be costly depending on sample sizes.

There a number of options to increase the accuracy of the impacts when samples sizes are smaller. One option is to increase the signal, the percent load reduction, relative to the background noise. Larger percent load are easier to detect with a higher degree of accuracy. Using a load shed strategy rather than reducing the duty cycle of AC units increases the reductions. So does relying on AC end-use data rather than whole household data. AC end-use data inherently has less noise and as a result, it is easier to identify the impact of load control, especially when temperatures are hot but not very hot. Another option is to rely on more sophisticated statistical analyses that explain of subtract some of the variation or background noise. One widely used statistical approach that is still transparent and easy to implement is a differences-in-differences calculation. It effectively calculates the bias during a set of non-event days such as non-event days with similar temperature conditions and nets it out of the demand reduction measurement calculation. This approaches increase the precision of the impact estimates at a low cost with relatively few barriers.

5.2 A Measurement Framework for Operations and Settlement

There are several advantages to using impact estimate tables for settlement. The tables provide an estimate of impacts in advance, settlement is fast and without uncertainty, settlement disputes are reduced, and it only requires one annual round of detailed analysis verified by the system operators (or its evaluation sub-contractor). Clearly, an impact estimate table needs to rely on estimated impacts from actual events. But it does split and simplify actual settlements from the task of re-calculating and verifying them.

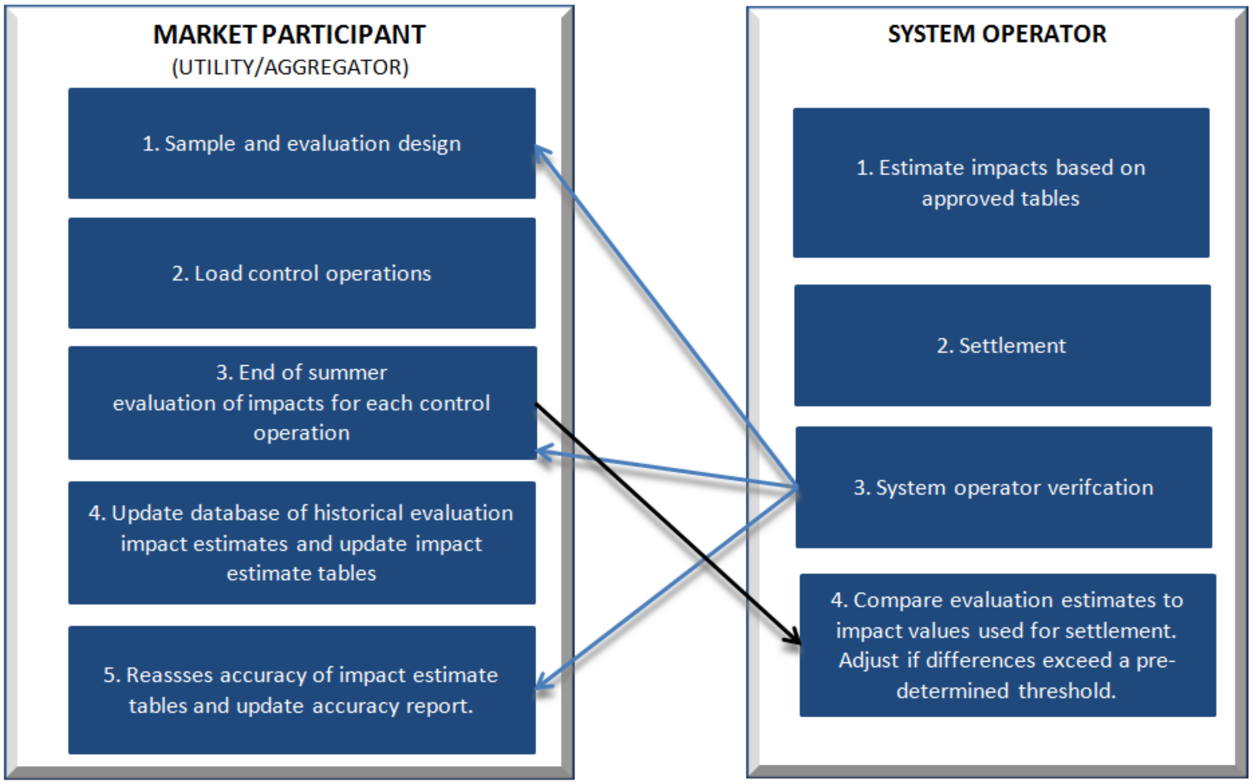
Figure 5-1 visually describes the proposed framework. In a short, it provides key junctures during which the accuracy of impact estimate tables and the data underlying them is verified. Otherwise, settlements are based on impact estimate tables that are agreed upon and verified in advance. From a market perspective, the accuracy of the settlement is the most important aspect. How the impact estimates are produces is a secondary component.

The framework calls for the development of impacts estimate tables on an annual basis and an assessment of their accuracy and consistency with evaluation impacts. The assessment of accuracy would be conducted much like it was done in this report by comparing predicted values to known simulated impacts. The accuracy assessment and underlying data would be made available for review by the system operators or its evaluation subcontractor. This allows system operators to verify the accuracy of the deemed savings in the impact estimate tables. In addition, a check for consistency of results is recommended. This involves comparing evaluation impacts at the end of the summer or year to the impact estimated from the tables and used for settlement. The results should agree. If they do not, it is an indicating that the impact estimate tables used in the prior summer were inaccurate. An option is to trigger adjustment to the settlement payments if the difference is greater than a pre-specified threshold – e.g., 10% of total payments.

One of the key constraints on smart meter data is the granularity of data collection, which is often limited to 15-minute or hourly interval in order to remain compatible with billing system and pre-existing data management systems. In contrast, settlement in real time markets is often at five minute increments. With ancillary services, the bulk of payments are tied to being on stand-by to deliver a resource with short notices and little ramp time. Only a small fraction of payments is associated with the few instances when spinning reserves are in fact dispatched. An option is for providers of AC load reduction to simply forego payments for electricity deliver and focus on stand-by payment and capacity payments.

While tables with pre-calculated load reductions per AC unit provide accurate estimates of demand reductions over course of the summer, they currently lack the precision needed for grid operations. Since the tables are based in part on annual evaluation results, they reflect any measurement error in those evaluations. Overall, when and where possible, it is highly recommended that evaluation impacts rely on random assignment of load control operations, large sample sizes and a weather matched difference-in-differences calculation. As discussed earlier, smart meters allow the use of larger sample sizes at lower costs and produce highly accurate impacts. Improving the measurement of the annual evaluation measurements should improve the accuracy of the tables. It is critical to continue expanding the body of empirical data used to develop the tables and, in particular, systematically include operations under different weather conditions and event hours. With these steps, it may be possible to reduce the uncertainty enough to utilize tables with pre-calculated load reductions per AC unit for operations.

Figure 5-1: Settlement Framework with Impact Estimate Tables (Deemed Savings)



5.3 Costs of settlement

Table 5-2 provides a range of estimated costs of settlement using AC end-use data, household data and a deemed savings approach relying on impact estimate tables. The key difference between each of the estimates is the data collection equipment, recruitment and installation costs. Collecting AC end-use data for settlement imposes substantial incremental costs. Data collection devices with remote data retrieval capabilities range from $300 to $1,500 per unit depending on the functionality. The estimates in Table 5-1 assume the needed functionality can be obtained at the lower end of the cost spectrum. In addition, it is necessary to factor in the cost of recruiting customers willing to have the devices installed. Acquisition and incentive cost are typically around $100 to $150 dollars per AC unit. Lastly, installing devices costs between $150-$250 per AC unit, depending on the complexity of the data collection device, AC units per location, 3-phase versus single phase power and other factors. These costs can add up quickly. To produce reliable estimates with AC data, the sample needs to be at least 500 units and preferably 1,000 units. Deploying the data collection system for 500 AC units can cost between $300,000 to

$500,000. The costs multiply if the impact estimates needs to be precise for specific geographic locations such as local capacity areas. If, for example, separate samples are required for the 16 load aggregation points in PG&E territory, this would expand costs to $4,800,000 to $8,000,000. While the data collections system can work for multiple years it would need to be updated to accurately changes in the population of load control participants. The additional costs associated with using AC end-use data for settlement are trivial compared to the data collection.

Using smart meter household data for settlement avoid most of those costs since the data collection systems are already in place. Deploying very large samples impose very small incremental costs in comparison to expanding sample sizes with AC end-use data. This allows for a high degree of precision highly localized impact estimates. The incremental data collection costs are limited to adjusting data collection intervals to read at 15-minute intervals and incremental data query costs. The primary cost is associated with the load reduction capability foregone by not dispatching a randomly selected group of customer during each load control operation. Other costs such as sample design, calculation of impacts and verification are similar to those with AC end-use data. There are two drivers for the range of uncertainty in the costs of impact and verification. One is the total hours to manage the data and update the calculations. The second source of uncertainty is the number of instances the program would be dispatched in order to help provide stabilize the electricity grid.

Table 5-2: Estimated Settlement Costs By Data Source

Component	AC End-use Data		Household Data		Impact Estimate Tables	
	Lower Bound	Upper Bound	Lower bound	Upper Bound	Lower bound	Upper Bound
Recommended Sample Size	500	1,000	1,000	2,000	Part of M&E	
Sample design and operations planning	$10,000	$20,000	$10,000	$20,000	Part of M&E	
Data collection - installation and equipments costs per AC unit	$600	$1,000	$5	$5	$5	$5*
Data transmittal	$10,000	$50,000	N/A	N/A	N/A	N/A
Calculation of Impacts	$40,000	$120,000	$40,000	$120,000	$20,000	$40,000
Verification Costs (ISO)	$20,000	$60,000	$20,000	$60,000	$15,000	$30,000
Update Impact Estimate Tables	N/A	N/A	N/A	N/A	$30,000	$60,000

Impact estimate tables do not pose substantial incremental costs if they rely on the same data collection, sample design and calculation of impacts generally used for routine measurement and evaluation of program impacts. In fact, the cost of calculating impacts is lower if they are aligned with the estimation of the evaluation impacts. It simplifies the settlement calculation to multiplying the relevant impact estimate from the table with the numbers of customer dispatched in each geographic location. It also lowers the system operator costs of verifying impacts since this is done once a year instead of for each and every load control operation. In exchange, there is and additional cost of producing the impact estimates tables and reassessing their accuracy once per year.

Appendix A. How and Why Does Baseline Accuracy Differ From Demand Reduction Accuracy

Both estimates of the counterfactual – what customers would have used in the absence of AC load curtailment – and demand reductions can contain error. While these errors are closely related, they are not one and the same. Importantly, different calculation methods compute demand reductions differently. The most common approach – day-matching or weather-matching baselines – calculate demand reduction as the difference between the estimate of the usage without curtailment operations and the actual loads during curtailments. On the other hand, regression models calculate demand reduction based on the regressions coefficients. The accuracy of regression results does not depend on how well the model explains electricity. As long as other factors are not confounded with the variables representing the curtailment operations, the regression models provide accurate estimates of the demand reduction. This is explained in more detailed below. With the third method, use of control groups, the demand reduction is calculated as the difference in load of customers that did and did not experience the AC curtailment.

A.1. Weather and Day-matching Baselines

Weather and day-matching baselines are within-in subject estimators. They rely on the customers own usage during a set of days where loads were not curtailed to infer what electricity use would have been absent curtailment operations – a baseline. Because these methods calculate demand reductions as the difference between the baseline and the actual loads during curtailments, errors in the baseline are magnified errors in the demand reduction estimates.

For example, if the actual demand reduction and true counterfactual – the electricity use absent curtailment – are 2 MW and 10 MW, a 5% upward bias in the settlement baseline will produce a baseline of 10.5 MW and a calculated load reduction of 2.5 MW (10.5 MW minus the metered load of 8 MW). While the baseline upward bias is 5%, the estimated demand reduction is biased upward by 25%; it is 2.5 MW rather than the actual 2.0 MW. If instead the true demand reduction were lower, say 1.0 MW (10%), a day-matching or weather-matching method would estimate a load reduction of 1.5 MW (10.5 MW minus the metered load of 9 MW). With the smaller load reduction, the 5% upward bias in the settlement baseline leads to a 50% upward bias in the estimated reductions – 1.5 MW of load reduction was estimated rather than the actual 1.0 MW. In general, relatively small baseline errors translate into larger errors in the estimated demand reductions.

Table A-1 summarizes how the percent error in demand reduction estimates varies as function of baseline error and the true percent demand reduction.

Table A-1: Relationship Between Load Reduction, Baseline Error and Impact Error

		Actual Percent Demand Reduction									
		5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
Baseline Percent Error	-10.0%	-200.0%	-100.0%	-66.7%	-50.0%	-40.0%	-33.3%	-28.6%	-25.0%	-22.2%	-20.0%
	-8.0%	-160.0%	-80.0%	-53.3%	-40.0%	-32.0%	-26.7%	-22.9%	-20.0%	-17.8%	-16.0%
	-6.0%	-120.0%	-60.0%	-40.0%	-30.0%	-24.0%	-20.0%	-17.1%	-15.0%	-13.3%	-12.0%
	-4.0%	-80.0%	-40.0%	-26.7%	-20.0%	-16.0%	-13.3%	-11.4%	-10.0%	-8.9%	-8.0%
	-2.0%	-40.0%	-20.0%	-13.3%	-10.0%	-8.0%	-6.7%	-5.7%	-5.0%	-4.4%	-4.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	2.0%	40.0%	20.0%	13.3%	10.0%	8.0%	6.7%	5.7%	5.0%	4.4%	4.0%
	4.0%	80.0%	40.0%	26.7%	20.0%	16.0%	13.3%	11.4%	10.0%	8.9%	8.0%
	8.0%	160.0%	80.0%	53.3%	40.0%	32.0%	26.7%	22.9%	20.0%	17.8%	16.0%
	10.0%	200.0%	100.0%	66.7%	50.0%	40.0%	33.3%	28.6%	25.0%	22.2%	20.0%

*The cells reflect the percent error in demand reduction estimates as a function of the percent error in the baseline and the true percent demand reduction.

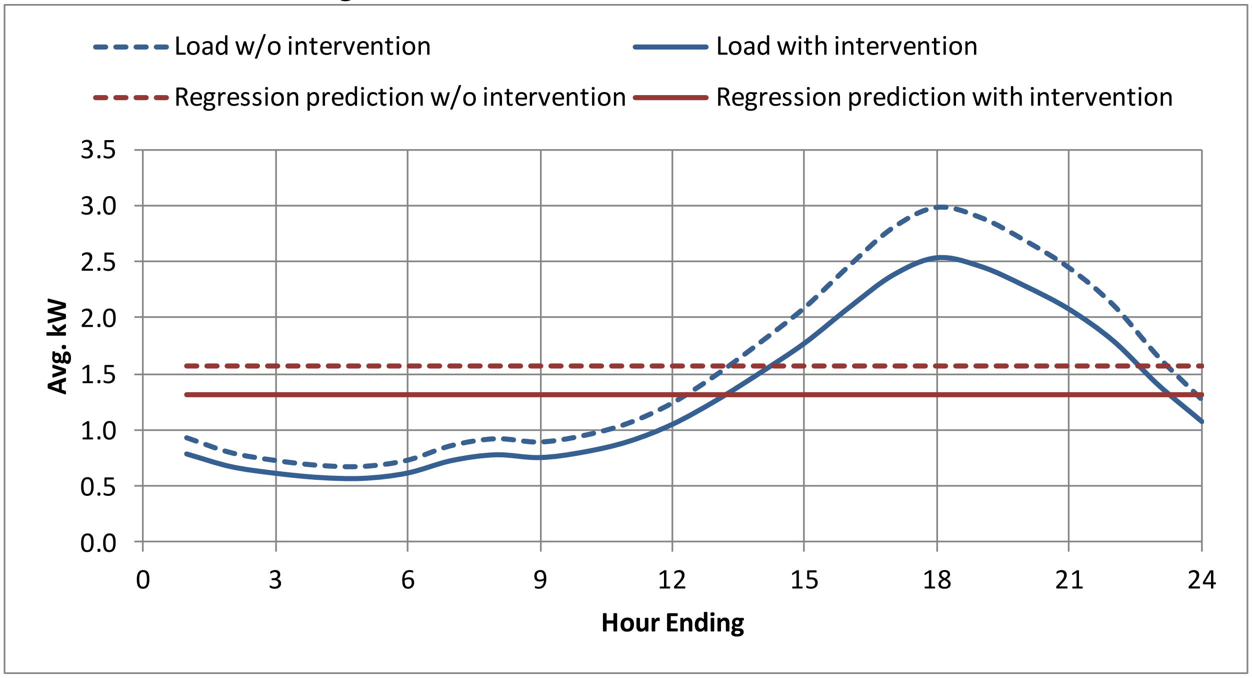
A.2. Regression Models

Regression models calculate demand reduction based on the regressions coefficients. Unlike day and weather-matching baselines, the accuracy of regression results do not depend on how well the regression model explains electricity. As long as other factors are not confounded with the variables representing the curtailment operations, the regression models provide accurate estimates of the demand reduction.

This point is counterintuitive at first and is best understood by considering an example were regression analysis is used to analyze and experiment. For illustrative purposes, we simulated an experiment where a set of 1,500 customers was randomly assigned an intervention while a second set of 1,500 customers acts as a control. Due to the random assignment, by definition other factors are not systematically related to the curtailments. For simplicity, the intervention reduced demand by 15% across all hours. Figure A-1 shows the actual loads with and without DR for the group that experienced the curtailment (which are known, because the experiment is simulated) and compares them to the regression predictions of loads with and without DR. The regression model used did not attempt to explain electricity use patterns. It simply modeled electricity use as a function of a constant and a variable indicating if the customer experienced the curtailment operation. Not surprisingly, the regression model does a poor job at predicting electricity use patterns. However, it calculates the demand reduction relatively well. The regression model estimates a reduction of 16.4%, which is relatively close to the true reduction of 15.0%.

In practice, a model that explains electricity use well typically minimizes the chance of confounding other factors with demand reductions. However, the main point of the illustration is that the accuracy of regression results, unlike the day and weather-matching method, is not linked to how well the regression model explains electricity use.

Figure A-1: Actual Loads With and Without DR

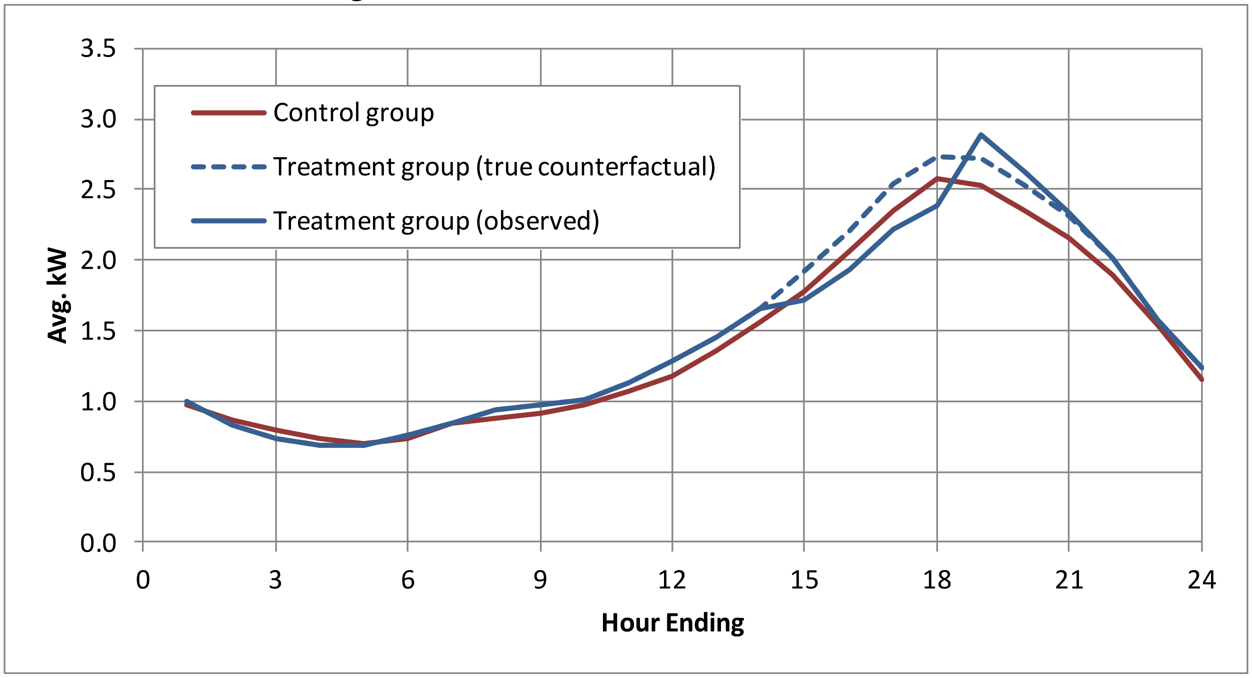


A.3. Control Groups

With control groups, demand reductions are estimated as the difference between a group that did not have their AC loads curtailed and one that did. The simplest way of calculating the demand reduction is to calculate the difference in electricity use between the two groups. In essence, the electricity use pattern of the group that was not curtailed is used to infer what electricity use patterns would have been if the curtailed group had not been curtailed. Ideally, the only systematic difference between the two groups is that one had their AC load curtailed and other group did not. A lack of systematic differences between the two groups eliminates alternative explanations besides the AC curtailment. The best way to ensure there are no systematic differences is to randomly assign customers to the curtailment and control groups and use large sample sizes.

As with day and weather-matching baselines, errors in the counterfactual – which in this case was provided by the control group – are magnified in the demand reduction estimate. Importantly, smaller sample sizes can introduce a substantial amount of error. Figure A-2 illustrates this point. The example is based on 300 customers randomly assigned to a curtailment group and another 300 randomly assigned to a control group. The differences between the two groups are entirely due to random variation in sampling. However, it is clear that control provides an inaccurate estimate of the true counterfactual. During the event hours from 2 PM to 6 PM the control underestimates the true counterfactual by 6.7%. However, because the demand reduction is estimated as the difference between the observed control and curtailment group loads, the impacts are underestimated by 54.4%. As with baseline methods, errors in the control group estimate of the counterfactual are magnified errors in the demand reduction estimate.

Figure A-2: Smaller Sizes Can Introduce Error



Appendix B. Feeder, Household and Air Conditioner End-use Data Sources

This appendix describes the feeder, household and AC end-use data sources used in the analysis. The various estimation methods were tested on feeder, household and AC end-use data representative of the PG&E SmartAC population. The choice of data source affects both the coverage across the service territory and the ability to produce impacts for localized areas, if needed for settlement. In each instance, a sample representative of residential AC units in the PG&E direct load control program, SmartAC was drawn. The intent was twofold: to assess the extent to which data was available across the territory and to ensure representative results. Rather than select the example feeders with high penetration levels, as had been done in prior studies, the goal was to assess whether the data sources could be widely used for program settlements.

B.1. Feeder Data

There are attractive features of feeder data from an operator's perspective. While not universal, the data collection system already tracks data in small increments, often times at intervals well below a minute. It meets a common system operator requirement for settlement at five-minute intervals. Another key advantage is that many feeders are able to transmit near real time data on electricity use. If there is a significant drop in feeder electricity load, the operators can see it and confirm it. The advantages of feeder data has less to do with the data itself and more to do with existing data collection systems and familiarity with them.

In total, 10% of the roughly 2,040 feeders with 1 or more customers enrolled in PG&E's AC control program were sampled. A proportional random sample was used to randomly select 10% of feeders across each climate region and feeder size strata identified. PG&E attempted to extract five-minute interval data for each feeder sampled.

Table B-1 summarizes the characteristics of the feeders sampled and those that did indeed have five-minute level data available. For clarity, all comparisons of the distribution across PG&E territory are at the AC unit level. Of the 204 feeders sampled, 85 (41.7%) of them tracked minute level feeder data. However, those same 85 feeders accounted for 54% of the control devices in 204 feeders initially sampled. The feeders with interval data available were generally larger than the average feeder. They both had more residential accounts than the population of feeders – 2,176 versus 1,811 accounts per feeder – and had a large number of accounts enrolled in the AC control program – 81 versus 62 per feeder. More importantly, a disproportionate share of the feeders with data is in the Greater Bay Area. In total, 63% of the AC units linked to feeders with data is in the Greater Bay Area while only 33% of all AC units in the program are in the area. This has several consequences, first the AC units in feeders with data experience milder weather and have less AC use than those in the population. This is evidenced by the difference in the weather sensitivity metric. Those accounts also have a smaller share of customers on the low income tariff than the broader SmartAC populations. Simply put, not only is the feeder data unavailable for a substantial share of the controllable AC units, but it is not representative of the program and severely under represents customers in hotter areas.

Table B-1: Comparison of SmartAC Feeder Population, Sample and Actual Available Data

Metric	SmartAC Population	Feeders Sampled	Actual Feeders Where Data was Available
Number of feeders	1998	204	85
Avg. residential accounts per feeder	1,811	1,812	2,176
Avg. SmartAC enrolled accounts per feeder	62	62	81
Total residential SmartAC accounts	124,047	12,451	6,713
Total Control Devices	140,934	14,146	7,627
2009 average monthly kWh	716	717	711
Weather sensitivity1	0.39	0.38	0.25
Low income rate (%)	27%	26%	15%
Greater Bay Area	33%	37%	63%
Greater Fresno	21%	22%	8%
Humboldt	0%	0%	0%
Kern	3%	2%	3%
Northern Coast	6%	6%	8%
Sierra	12%	11%	6%
Stockton	10%	10%	0%
Other	16%	12%	13%

[1] Weather sensitivity was calculated as the correlation between monthly usage (kWh) and total CDH (Base 70°F) over the same time period.

B.2. Household Data

Until relatively recently, household data on hourly and sub-hourly electricity use was not available for most customers and was expensive to collect. When it was available, it usually was for a sample of customers representative of the general population. Substantial changes in technology, and in particular the adoption of smart meters, have made settlement with household data feasible. As of November 2010, over 82% of the 125,000 SmartAC participants had smart meters in place and nearly all residential customers will have smart meters by the end of 2011. In total, 70% of SmartAC participants had smart meter data available throughout the summer of 2010.

The smart meter household data was drawn from the same 204 feeders sampled for which feeder data was sought in order to allow direct comparisons. For each of the 204 feeders, 100 households in the AC control program were sampled if available. If enrollment in a feeder was lower, the full population was included in the analysis. The smart meters in each feeder were weighted for the overall number of feeder households enrolled in the AC control program. For example, if a feeder had 450 participating households, each of the 100 randomly sampled in the feeder was assigned a weight of 4.5, since they each represent 4.5 households. In another feeder with 68 participating households, all of the customers with smart meters were sampled and assigned a weight of 1.0.

While coverage by smart meters will be nearly universal in PG&E territory within a year, not all of them had smart meters in place by the summer of 2010, mainly because the installation of those meters was still underway. Table B-2 compares the SmartAC population to the feeders sampled, to the actual data obtained and weighted. Unlike feeder data, it was known in advance that 8,800 of the 12,450 accounts had smart meters in place throughout the 2010 summer. In total, using the sampling strategy described in the above paragraph, data for 6,010 households was requested and data for 5,988 households (99.6%) of the data was delivered. While only 132 of 204 feeders sampled had customers with smart meters in place, the meters were already located where it counted and provided representation for 88.5% (12,530/14,146) of the AC units. This is due to how the smart meters were deployed in the PG&E territory. Meter installations were first deployed in warmer parts of the territory, were AC penetration is higher and, by connection, participation in SmartAC is higher. The meter installations have also covered most urban and suburban areas, where feeders generally have a larger number of connected accounts.

Table B-2: Comparison of SmartAC Population, Feeders Sampled and Household Data Sampled

Metric	SmartAC Population	Feeders Sampled	Household Data Sampled (No Weights)	Household Data Sampled (Weights)
Number of feeders	1998	204	132	132
Avg. residential accounts per feeder	1,811	1,812	2,176	2,176
Avg. SmartAC accounts per feeder	62	62	81	81
Residential SmartAC accounts	124,047	12,451	5,988	11,029
Control devices	140,934	14,146	6,803	12,530
2009 average monthly kWh	716	717	725	726
Weather sensitivity[1]	0.39	0.38	0.43	0.39
Low income rate (%)	27%	26%	30%	26%
Greater Bay Area	33%	37%	29%	37%
Greater Fresno	21%	22%	32%	25%
Humboldt	0%	0%	0%	0%
Kern	3%	2%	4%	2%
Northern Coast	6%	6%	5%	5%
Sierra	12%	11%	9%	10%
Stockton	10%	10%	9%	12%
Other	16%	12%	13%	10%

[1] Weather sensitivity was calculated as the correlation between monthly usage (kWh) and total CDH (Base 70°F) over the same time period.

As noted earlier, by default, PG&E stores residential smart meter data at hourly intervals and can adjust the interval of data collection down to 15-minute intervals without requiring major modifications to billing systems.

B.3. Air Conditioner End-use Data

Unlike feeder or household data, AC end-use data is not available without special effort to sample AC units, install data collection devices and retrieve them. This has two implications. First, coverage with AC units is limited and requires the use of samples. Second, the cost per unit sampled is substantially higher and depending on the technology capabilities can impose high data retrieval costs.

PG&E collected AC end-use data at five-minute intervals or less in 2008, 2009 and 2010, primarily to evaluate impacts at the program level. In 2008, the residential sample included roughly 700 AC units and 20 experimental operations lasting 4 to 6 hours each were called to better understand load reduction capabilities. In 2009, three different residential samples of AC units were employed. The first was designed to reflect the SmartAC population and consisted of 547 AC units. The sample was stratified by three climate regions, household vintage and intentionally sampled areas of the Central Valley. The AC units for this sample were not controlled. The second 2009 sample targeted AC load of customers that were not enrolled in the SmartAC program in order to assess whether participants used their ACs differently. The third sample consisted of 500 AC units concentrated on 4 feeders and was designed to assess the potential for using AC load control for spinning reserves. As noted earlier, the 2,000 AC units in those 4 feeders were instructed to fully shed load nearly 70 times for 15 minutes at a time. In 2010, the residential end-use sample consisted of 330 AC units. Those units were called for 14 control events lasting 4 hours each.

To test settlement alternatives, the 2009 primary program sample consisting of 547 AC units was used. Because no events were called, the sample reflects how customers enrolled in SmartAC naturally operate their AC units. It contains unperturbed AC use data. In contrast, the AC units in the 2008 and 2010 samples were controlled on a large number of days, particularly when it was hot. Because AC end-use data is not universal and required using a sample, it cannot provide information for specific feeders as was the case for the feeder and household data.

Appendix C. Mathematical Expression of Regression Models

No.	Model Description	Mathematical Expression
1	*Treatment variables and no day or hourly lags or leads.* This model estimates demand reductions as a function of the temperature during the hour, as measured by cooling degree hours (CDH), and total heat intensity in the day prior to curtailment (past24hrCDH). Several additional variables are included to explain variation in electricity use so the demand reduction signal can be better detected, including: ▪ The effect of hour, weekend, and school vacations ▪ Total heat intensity in the day prior ▪ Heat intensity on during each time period ▪ The interaction between school vacation periods and the use of cooling. The model can be applied for both ex post estimation and to forecast available AC loads and demand reduction potential. It uses data from prior curtailments to inform the estimate of the demand reduction for the curtailment in question.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past24hrCDH_{i,t} + \beta_{i,h} \cdot past24hrCDH_{i,t}^2 + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH_{i,t}^2 + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \varepsilon_{i,h,t}$
2	*Treatment variables with a hourly lag.* This model estimates demand reductions as a function of the temperature during the hour, as measured by cooling degree hours (CDH), and total heat intensity in the day prior to curtailment (past24hrCDH). The same additional variables as in Model 1 are included to explain variation in electricity use so the demand reduction signal can be better detected. In addition, this model includes the electricity use two hours prior to the time period in question to explain variation in electricity use. The model can be applied for ex post estimation. It requires continuous data uploading to use it for forecasting available AC loads and demand reduction potential for operations.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past24hrCDH_{i,t} + \beta_{i,h} \cdot past24hrCDH_{i,t}^2 + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH_{i,t}^2 + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \beta_{i,h} \cdot kw_{i,h-2,t} + \varepsilon_{i,h,t}$

No.	Model Description	Mathematical Expression
3	*Treatment variables with hourly lags and leads.* Like Models 1 and 2, this model estimates demand reductions as a function of the temperature during the hour, as measured by cooling degree hours (CDH), and total heat intensity in the day prior to curtailment (past24hrCDH). The same additional variables as in Model 1 are included to explain variation in electricity use so the demand reduction signal can be better detected. In addition, this model includes the electricity use in both hours preceding and after the time period in question to explain variation in electricity use. The model can be applied for ex post estimation, but cannot be used to forecast available AC loads and demand reduction potential for operations.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past24hrCDH_{i,t} + \beta_{i,h} \cdot past24hrCDH_{i,t}^2 + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH_{i,t}^2 + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \sum_{j=-3,-2,2,3} \beta_{i,h} \cdot kw_{i,h+j,t} + \varepsilon_{i,h,t}$
4	*No treatment variables but use of hourly lags and leads.* This model differs from the prior three models in that the demand reductions is not explicitly calculated using regression coefficient. Rather the regression is used to provide an estimate of load without curtailment and the impacts are calculated as the difference between the reference load and the metered load during the curtailment. Except for the lack of treatment variables, the explanatory variables are the same as in Model 3.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past24hrCDH_{i,t} + \beta_{i,h} \cdot past24hrCDH_{i,t}^2 + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH_{i,t}^2 + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \sum_{j=-2,-1,2,3} \beta_{i,h} \cdot kw_{i,h+j,t} + \varepsilon_{i,h,t}$

Appendix D. Example of Same-day Adjustment Calculation

A same-day adjustment is a way to reduce the error between the baseline and actual loads during an event period. The idea behind the adjustment is that the period before the event – when customers' loads are unperturbed – can be used to check the accuracy of the baseline calculation for that day and adjust the baseline up or down to make it more accurate. If the baseline calculation is below the actual load before the event, then it is assumed that the baseline would also be below what the unperturbed load would have been during the event period. Similarly, if the baseline is above the actual load before the event, it is also assumed to be above what the unperturbed load would have been during the event period. *In other words, it is assumed that changes in pre-event are not due to load shifting to those hours or gaming.* This relationship between the two loads is used to determine the same-day adjustment for that customer. The calculation is illustrated using a concrete example:

Suppose a 3-hour same-day adjustment is to be applied to a customer during an event window of 3 PM to 7 PM. The hours 11 AM to 2 PM are used to determine the adjustment (this includes a one-hour buffer between the adjustment period and the event period – this buffer is explained below).

Suppose a customer has the actual hourly load shown in Table D-1 below. For simplicity of the example, only unperturbed load is considered – that is, the customer's use is known in the absence of the event. This allows the focus to be on the effect of the adjustment and how the adjusted baseline compares to the number it is supposed to predict. Suppose the 10-of-10 baseline method with no adjustment is used for this customer giving the baseline also shown in Table D-1. Comparing across loads during the same hour, it is obvious that the unadjusted baseline is higher than the actual – both before the event and during the event. The baseline therefore over-predicts the actual load.

Table D-1: Actual Use and Unadjusted Top 10-of-10 Baseline for a Particular Customer

Hour Starting	10:00 AM	11:00 AM	12:00 PM	1:00 PM	2:00 PM	3:00 PM	4:00 PM	5:00 PM	6:00 PM	7:00 PM
Actual Unperturbed Use (kW)	4.2	4.0	3.7	4.1	4.0	4.1	4.0	4.1	3.9	3.9
10 of 10 Baseline with No Adjustment (kW)	5.1	5.0	4.8	5.1	4.9	5.0	5.1	5.3	4.9	5.0

The same-day adjustment consists of applying a multiplier to the baseline during the event period. To calculate that multiplier two numbers must be calculated: the average actual use during the pre-event period 11 AM to 2 PM and the average baseline with no adjustment during the same period. In this example, the average actual use during the period is:

$$avguse_{pre-event} = \frac{4.0 + 3.7 + 4.1}{3} = 3.93\ kW$$

The average baseline with no adjustment is:

$$avgbaseline_{pre-event} = \frac{55.0 + 4.8 + 5.1}{3} = 4.97\ kW$$

The multiplier to be applied is the ratio of the average actual use to the average unadjusted baseline:

$$actual - to - baselineloadratio = \frac{avguse_{pre-event}}{avgbaseline_{pre-event}} = \frac{3.93\ kW}{4.97\ kW} = 0.79$$

The rationale behind applying this multiplier to the baseline during the event period is that given the proximity of the pre-event period to the event window, the baseline percentage error would be roughly the same during the event window as it was during the pre-event period. Again, this assumes customers are not shifting load to pre-event hours. To adjust the baseline closer to what the actual load would be without an event, the baseline of each event hour is multiplied by the actual-to-baseline load ratio. In this case, that ratio is 0.79.

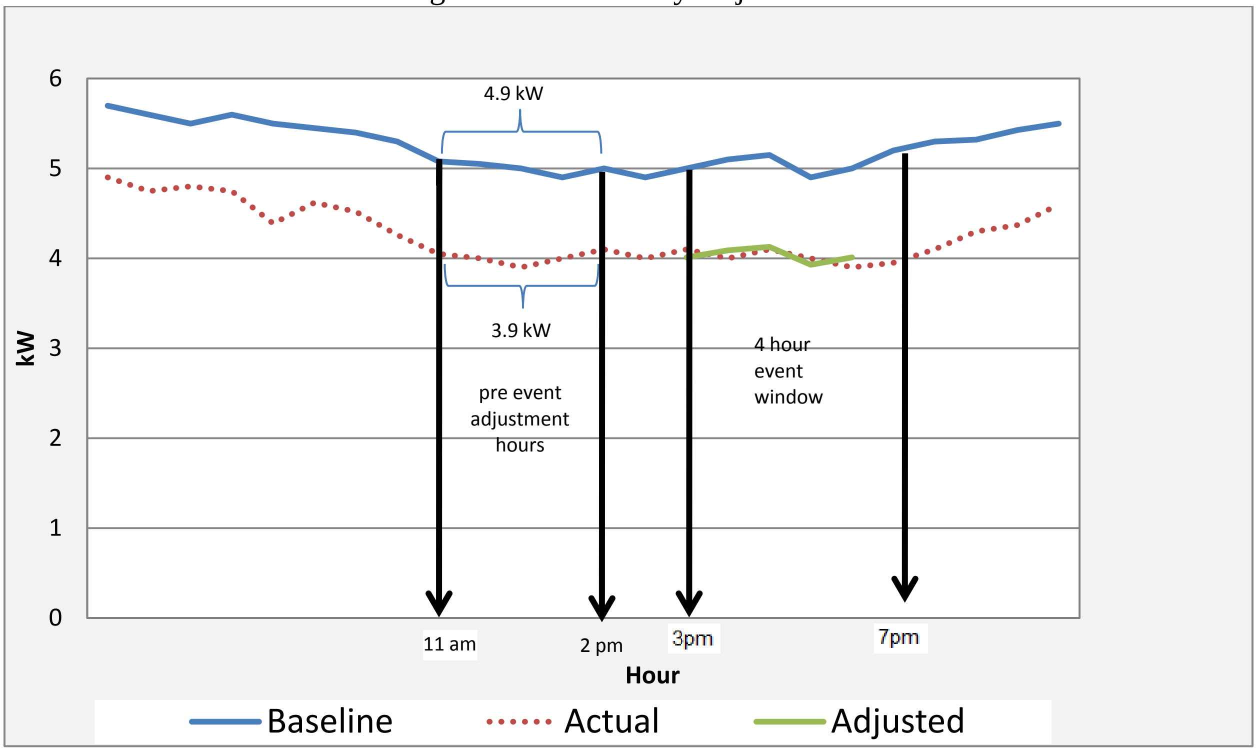
The result of this calculation for this example is shown in Table D-2. The numbers in the adjusted row equal the numbers in the unadjusted row, multiplied by 0.8. The adjusted baseline numbers are much closer to the actual use.

Table D-2: Unadjusted 10-of-10 Baseline and 10-of-10 Baseline with Four-hour Same-day Adjustment for a Particular Customer

Hour Starting	3:00 PM	4:00 PM	5:00 PM	6:00 PM
Top 10-of-10 baseline with no adjustment (kW)	5.0	5.1	5.3	4.9
Top 10-of-10 baseline with same-day adjustment (kW)	4.0	4.0	4.2	3.9
Actual Unperturbed Use (kW)	4.1	4.0	4.1	3.9

Figure D-1 shows the same example graphically. The black arrows identify the 4-hour event period lasting from 3 PM to 7 PM and the 3-hour pre-event period from 11 AM to 2 PM (with a buffer from 2 PM to 3 PM). The pre-event period baseline average is 5 kW, while the pre-event actual use is 4 kW. This results in an actual-to-baseline load ratio of 0.8. Multiplying the baseline load during each event hour by the ratio gives the adjusted baseline. Notice how the adjusted load is much closer to the unperturbed actual load during the event hours. The error between the baseline and actual loads has been reduced dramatically. This reduces load impact errors and generates more accurate settlement payments. Figure D-1 shows the effect baseline adjustment has on load impact errors.

Figure D-1: Same-day Adjustment



Assume the contracted reduction for the customer in this example is 3 kW. However, the unperturbed actual load of the customer is only about 3.9 kW. Therefore, this customer needs to reduce load by only 2 kW to achieve full compliance with settlement rules that do not require in-day adjustment. This equates to a load impact error percentage of roughly 33%. The utility would only receive 66% of the reduction for which it has paid. Same-day adjustments, by bringing the baseline closer to the unperturbed actual load, can dramatically reduce this error and ensure utilities receive a greater share of the expected reduction. The reverse is also true. Some customers have to deliver more than the contracted demand reduction to comply with settlement rules. They essentially deliver more demand reduction than they are paid for. In-day adjustments generally reduce the degree of over and under payments to individual settlement accounts.

Appendix E. Average Air Conditioner Demand by Climate Region, Heat Intensity and Hour

Table E-1: Very Hot Climate Region (Central Valley - Fresno/Bakersfield)

Hour Ending	0 CDD	1-2 CDD	3-4 CDD	5-6 CDD	7-8 CDD	9-10 CDD	11-12 CDD	13-14 CDD	15-16 CDD	17-18 CDD	19-20 CDD	21-22 CDD	23-24 CDD	24 CDD or more
1:00	0.01	0.03	0.04	0.05	0.05	0.08	0.06	0.09	0.11	0.15	0.17	0.22	0.26	0.31
2:00	0.01	0.01	0.02	0.03	0.03	0.05	0.05	0.06	0.07	0.10	0.12	0.15	0.18	0.22
3:00	0.01	0.01	0.02	0.02	0.01	0.03	0.03	0.04	0.04	0.06	0.08	0.11	0.12	0.16
4:00	0.01	0.01	0.01	0.02	0.01	0.03	0.03	0.03	0.02	0.05	0.06	0.07	0.09	0.12
5:00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.04	0.05	0.06	0.09
6:00	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.03	0.04	0.05	0.06	0.08
7:00	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.04	0.04	0.06	0.09
8:00	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.02	0.04	0.05	0.06	0.10
9:00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.05	0.06	0.09	0.16
10:00	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.03	0.03	0.05	0.07	0.11	0.15	0.24
11:00	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.06	0.09	0.14	0.23	0.26	0.39
12:00	0.02	0.02	0.01	0.04	0.04	0.04	0.08	0.08	0.11	0.18	0.29	0.39	0.43	0.59
13:00	0.03	0.01	0.02	0.05	0.06	0.09	0.12	0.16	0.24	0.33	0.48	0.63	0.70	0.82
14:00	0.02	0.01	0.02	0.07	0.09	0.16	0.18	0.28	0.43	0.55	0.71	0.92	1.00	1.06
15:00	0.02	0.02	0.05	0.09	0.14	0.24	0.30	0.43	0.66	0.82	0.96	1.18	1.26	1.29
16:00	0.02	0.04	0.05	0.15	0.20	0.32	0.41	0.63	0.90	1.06	1.19	1.44	1.44	1.48
17:00	0.02	0.07	0.06	0.21	0.26	0.42	0.52	0.78	1.06	1.24	1.35	1.62	1.67	1.56
18:00	0.02	0.08	0.06	0.25	0.31	0.49	0.58	0.85	1.19	1.29	1.42	1.65	1.68	1.58
19:00	0.02	0.09	0.07	0.23	0.30	0.42	0.54	0.76	1.09	1.23	1.32	1.54	1.60	1.54
20:00	0.01	0.04	0.07	0.22	0.22	0.35	0.40	0.64	0.91	1.02	1.09	1.31	1.43	1.37
21:00	0.03	0.03	0.06	0.12	0.16	0.28	0.32	0.51	0.72	0.78	0.87	1.04	1.16	1.12
22:00	0.02	0.02	0.05	0.08	0.12	0.21	0.23	0.34	0.51	0.57	0.63	0.77	0.88	0.88
23:00	0.03	0.03	0.03	0.10	0.11	0.16	0.17	0.26	0.31	0.39	0.43	0.52	0.62	0.62
0:00	0.02	0.09	0.01	0.05	0.07	0.08	0.10	0.15	0.21	0.25	0.27	0.34	0.43	0.42

Table E-2: Hot Climate Region (Central Valley - Sacramento/Stockton/Fairfield)

Hour Ending	0 CDD	1-2 CDD	3-4 CDD	5-6 CDD	7-8 CDD	9-10 CDD	11-12 CDD	13-14 CDD	15-16 CDD	17-18 CDD	19-20 CDD	21-22 CDD	23-24 CDD
1:00	0.02	0.09	0.09	0.06	0.12	0.11	0.11	0.16	0.20	0.21	0.23	0.26	0.11
2:00	0.02	0.06	0.06	0.04	0.08	0.08	0.07	0.12	0.15	0.15	0.16	0.18	0.17
3:00	0.01	0.04	0.04	0.03	0.06	0.05	0.05	0.08	0.12	0.12	0.12	0.13	0.15
4:00	0.01	0.03	0.04	0.02	0.05	0.04	0.04	0.06	0.08	0.09	0.08	0.10	0.13
5:00	0.01	0.02	0.02	0.02	0.04	0.04	0.03	0.05	0.06	0.06	0.07	0.08	0.09
6:00	0.00	0.02	0.02	0.01	0.03	0.03	0.02	0.04	0.05	0.05	0.08	0.07	0.06
7:00	0.00	0.01	0.01	0.01	0.02	0.02	0.01	0.03	0.04	0.03	0.08	0.06	0.13
8:00	0.01	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.05	0.05	0.13	0.09	0.18
9:00	0.01	0.03	0.02	0.02	0.03	0.03	0.03	0.04	0.07	0.06	0.14	0.13	0.30
10:00	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.06	0.13	0.11	0.18	0.22	0.65
11:00	0.02	0.06	0.06	0.05	0.07	0.08	0.08	0.10	0.22	0.18	0.28	0.35	0.77
12:00	0.02	0.08	0.10	0.08	0.10	0.12	0.14	0.17	0.35	0.34	0.41	0.51	1.14
13:00	0.04	0.11	0.13	0.12	0.16	0.18	0.22	0.29	0.50	0.55	0.60	0.67	1.34
14:00	0.05	0.16	0.17	0.18	0.26	0.29	0.36	0.43	0.70	0.72	0.88	0.91	1.51
15:00	0.07	0.22	0.24	0.24	0.37	0.40	0.53	0.59	0.91	1.04	1.24	1.19	1.62
16:00	0.08	0.30	0.30	0.33	0.50	0.56	0.72	0.81	1.17	1.32	1.58	1.37	1.71
17:00	0.12	0.41	0.39	0.45	0.65	0.76	0.97	1.01	1.38	1.54	1.81	1.48	1.69
18:00	0.14	0.44	0.43	0.47	0.73	0.84	1.10	1.13	1.43	1.61	1.82	1.51	1.74
19:00	0.14	0.40	0.39	0.48	0.69	0.83	1.12	1.13	1.39	1.48	1.89	1.45	1.65
20:00	0.11	0.33	0.27	0.38	0.52	0.66	0.93	0.95	1.16	1.24	1.54	1.23	1.45
21:00	0.09	0.22	0.19	0.26	0.35	0.48	0.69	0.72	0.89	0.98	1.25	1.05	0.86
22:00	0.06	0.17	0.15	0.18	0.28	0.35	0.49	0.54	0.68	0.76	0.91	0.87	0.73
23:00	0.04	0.12	0.10	0.11	0.20	0.22	0.35	0.35	0.46	0.50	0.61	0.64	0.55
0:00	0.04	0.09	0.07	0.09	0.15	0.15	0.22	0.23	0.35	0.36	0.43	0.46	0.38

Table E-3: Warm Climate Region (Bay Area - Diablo Valley / San Jose)

Hour Ending	0 CDD	1-2 CDD	3-4 CDD	5-6 CDD	7-8 CDD	9-10 CDD	11-12 CDD	13-14 CDD	15-16 CDD	17-18 CDD	19-20 CDD	21-22 CDD	23-24 CDD
1:00	0.02	0.03	0.05	0.03	0.06	0.06	0.09	0.09	0.12	0.11	0.19	0.16	0.24
2:00	0.01	0.02	0.04	0.02	0.03	0.03	0.07	0.09	0.09	0.07	0.12	0.16	0.25
3:00	0.01	0.01	0.02	0.01	0.01	0.02	0.06	0.05	0.05	0.05	0.08	0.15	0.11
4:00	0.00	0.00	0.02	0.01	0.01	0.01	0.05	0.04	0.04	0.03	0.03	0.15	0.08
5:00	0.00	0.00	0.01	0.01	0.00	0.00	0.04	0.03	0.02	0.03	0.02	0.13	0.04
6:00	0.00	0.00	0.01	0.01	0.00	0.00	0.03	0.04	0.02	0.03	0.01	0.08	0.02
7:00	0.00	0.00	0.01	0.01	0.00	0.00	0.03	0.03	0.02	0.03	0.03	0.07	0.01
8:00	0.00	0.00	0.01	0.01	0.00	0.00	0.03	0.04	0.02	0.03	0.02	0.10	0.04
9:00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.04	0.02	0.03	0.01	0.06	0.06
10:00	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.04	0.02	0.03	0.03	0.07	0.09
11:00	0.00	0.00	0.01	0.01	0.01	0.01	0.04	0.05	0.06	0.06	0.09	0.16	0.19
12:00	0.00	0.01	0.01	0.02	0.01	0.01	0.06	0.07	0.12	0.14	0.18	0.36	0.49
13:00	0.00	0.01	0.03	0.03	0.02	0.03	0.13	0.14	0.24	0.26	0.48	0.69	0.82
14:00	0.01	0.02	0.04	0.05	0.06	0.07	0.23	0.22	0.39	0.42	0.81	1.02	1.35
15:00	0.01	0.05	0.08	0.10	0.13	0.17	0.34	0.45	0.63	0.67	1.16	1.25	1.64
16:00	0.03	0.08	0.12	0.19	0.25	0.29	0.50	0.68	0.93	1.01	1.41	1.72	1.87
17:00	0.05	0.13	0.17	0.30	0.36	0.45	0.71	0.84	1.13	1.24	1.53	1.95	1.91
18:00	0.05	0.15	0.19	0.35	0.47	0.58	0.84	0.98	1.29	1.32	1.66	2.01	2.35
19:00	0.05	0.15	0.18	0.35	0.47	0.59	0.86	1.04	1.23	1.35	1.65	1.86	2.31
20:00	0.04	0.11	0.14	0.26	0.34	0.43	0.72	0.89	0.96	1.10	1.54	1.67	2.11
21:00	0.03	0.08	0.11	0.17	0.21	0.29	0.52	0.64	0.67	0.86	1.16	0.92	1.77
22:00	0.02	0.05	0.08	0.10	0.16	0.19	0.33	0.44	0.47	0.64	0.67	0.75	1.02
23:00	0.02	0.04	0.06	0.06	0.12	0.13	0.23	0.29	0.29	0.45	0.40	0.55	0.59
0:00	0.01	0.03	0.04	0.05	0.06	0.08	0.15	0.18	0.18	0.25	0.28	0.36	0.28

Appendix F. Table of Demand Reductions per AC Unit by Climate Region, Heat Intensity and Hour

Table F-1: Very Hot Climate Region (Central Valley - Fresno/Bakersfield) – 50% AC cycling

Hour Ending	0 CDD	1-2 CDD	3-4 CDD	5-6 CDD	7-8 CDD	9-10 CDD	11-12 CDD	13-14 CDD	15-16 CDD	17-18 CDD	19-20 CDD	21-22 CDD	23-24 CDD	24 CDD or more
1:00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.06	0.07
2:00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.05
3:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.04
4:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02
5:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02
6:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02
7:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02
8:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02
9:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03
10:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.05
11:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.05	0.06	0.09
12:00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.04	0.06	0.09	0.10	0.15
13:00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.04	0.05	0.07	0.11	0.15	0.17	0.21
14:00	0.00	0.00	0.00	0.01	0.02	0.03	0.04	0.06	0.09	0.13	0.17	0.23	0.25	0.30
15:00	0.00	0.00	0.01	0.01	0.02	0.04	0.06	0.09	0.15	0.20	0.24	0.31	0.34	0.38
16:00	0.00	0.01	0.01	0.02	0.04	0.06	0.09	0.14	0.21	0.26	0.31	0.39	0.40	0.44
17:00	0.00	0.01	0.01	0.03	0.05	0.08	0.11	0.18	0.26	0.31	0.36	0.45	0.47	0.46
18:00	0.00	0.02	0.01	0.04	0.06	0.09	0.13	0.20	0.30	0.33	0.38	0.46	0.48	0.47
19:00	0.00	0.01	0.01	0.04	0.06	0.08	0.11	0.17	0.27	0.31	0.35	0.42	0.45	0.46
20:00	0.00	0.01	0.01	0.03	0.04	0.06	0.08	0.14	0.22	0.25	0.27	0.34	0.39	0.39
21:00	0.00	0.01	0.01	0.02	0.03	0.05	0.06	0.11	0.16	0.19	0.21	0.26	0.31	0.31
22:00	0.00	0.00	0.01	0.01	0.02	0.03	0.04	0.07	0.11	0.13	0.15	0.19	0.22	0.23
23:00	0.00	0.01	0.00	0.01	0.02	0.02	0.03	0.05	0.06	0.08	0.10	0.12	0.15	0.16
0:00	0.00	0.02	0.00	0.01	0.01	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.10	0.10

Table F-2: Hot Climate Region (Central Valley - Sacramento/Stockton/Fairfield) – 50% AC Cycling

Hour Ending	0 CDD	1-2 CDD	3-4 CDD	5-6 CDD	7-8 CDD	9-10 CDD	11-12 CDD	13-14 CDD	15-16 CDD	17-18 CDD	19-20 CDD	21-22 CDD	23 or more CDD
1:00	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.02
2:00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04
3:00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03
4:00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.03
5:00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.02
6:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
7:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02
8:00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.03
9:00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.04
10:00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.04	0.04	0.14
11:00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.04	0.03	0.06	0.07	0.17
12:00	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.07	0.07	0.10	0.12	0.27
13:00	0.00	0.01	0.02	0.02	0.03	0.04	0.04	0.06	0.11	0.12	0.15	0.17	0.34
14:00	0.00	0.02	0.03	0.03	0.05	0.06	0.08	0.10	0.17	0.17	0.23	0.24	0.40
15:00	0.01	0.03	0.04	0.04	0.08	0.09	0.12	0.14	0.22	0.26	0.34	0.33	0.47
16:00	0.01	0.05	0.05	0.07	0.11	0.12	0.17	0.20	0.30	0.36	0.44	0.39	0.53
17:00	0.01	0.07	0.07	0.09	0.14	0.18	0.24	0.25	0.35	0.42	0.51	0.43	0.53
18:00	0.01	0.07	0.07	0.09	0.15	0.19	0.27	0.29	0.37	0.44	0.51	0.44	0.54
19:00	0.01	0.06	0.06	0.09	0.14	0.18	0.26	0.28	0.35	0.38	0.51	0.41	0.49
20:00	0.01	0.05	0.04	0.07	0.10	0.13	0.20	0.21	0.27	0.30	0.40	0.33	0.41
21:00	0.01	0.03	0.03	0.04	0.06	0.09	0.13	0.15	0.18	0.22	0.29	0.25	0.19
22:00	0.01	0.02	0.02	0.03	0.04	0.06	0.09	0.10	0.13	0.15	0.19	0.19	0.13
23:00	0.01	0.02	0.02	0.02	0.03	0.03	0.06	0.06	0.08	0.09	0.12	0.13	0.09
0:00	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.06	0.08	0.09	0.05

Table F-3: Warm Climate Region (Bay Area - Diablo Valley / San Jose) – 50% AC Cycling

Hour Ending	0 CDD	1-2 CDD	3-4 CDD	5-6 CDD	7-8 CDD	9-10 CDD	11-12 CDD	13-14 CDD	15-16 CDD	17-18 CDD	19-20 CDD	21-22 CDD	23 or more CDD
1:00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.04
2:00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.04
3:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02
4:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.01
5:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.01
6:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00
7:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00
8:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.02	0.01
9:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.02
10:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.02	0.03
11:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.05	0.06
12:00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.10	0.13
13:00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.05	0.05	0.10	0.19	0.20
14:00	0.00	0.00	0.00	0.01	0.01	0.01	0.04	0.04	0.09	0.09	0.19	0.29	0.35
15:00	0.00	0.01	0.01	0.01	0.02	0.03	0.06	0.10	0.15	0.16	0.29	0.37	0.47
16:00	0.00	0.01	0.02	0.03	0.05	0.06	0.11	0.16	0.25	0.27	0.37	0.50	0.55
17:00	0.00	0.02	0.02	0.05	0.08	0.10	0.17	0.21	0.30	0.33	0.41	0.57	0.54
18:00	0.00	0.02	0.03	0.06	0.10	0.13	0.19	0.24	0.33	0.34	0.43	0.56	0.67
19:00	0.00	0.02	0.02	0.06	0.09	0.13	0.19	0.24	0.31	0.34	0.42	0.51	0.62
20:00	0.00	0.01	0.02	0.04	0.06	0.09	0.15	0.19	0.22	0.25	0.38	0.41	0.54
21:00	0.00	0.01	0.01	0.02	0.03	0.05	0.10	0.13	0.14	0.19	0.26	0.21	0.42
22:00	0.00	0.01	0.01	0.01	0.02	0.03	0.05	0.08	0.09	0.13	0.14	0.16	0.24
23:00	0.00	0.01	0.01	0.01	0.02	0.02	0.04	0.05	0.05	0.09	0.07	0.11	0.13
0:00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.03	0.05	0.05	0.06	0.05

Appendix G. Effect on Sampling Error on Within Subject Calculation Methods

Table G-1: Demand Reduction Margin of Error as a Function of Sampling and Estimation Error For Within-subject Calculation Methods

Estimation Margin of Error (95% Confidence)	Sampling Margin of Error (95% confidence)									
	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%
2%	2.8%	4.5%	6.3%	8.2%	10.2%	12.2%	14.2%	16.2%	18.1%	20.2%
4%	4.5%	5.7%	7.3%	8.9%	10.8%	12.6%	14.6%	16.5%	18.5%	20.4%
6%	6.3%	7.2%	8.5%	10.0%	11.7%	13.4%	15.3%	17.1%	18.8%	20.9%
8%	8.3%	9.0%	10.0%	11.3%	12.9%	14.5%	16.1%	18.0%	19.7%	21.5%
10%	10.3%	10.7%	11.6%	12.9%	14.3%	15.6%	17.2%	18.9%	20.6%	22.3%
12%	12.2%	12.5%	13.5%	14.3%	15.8%	17.0%	18.4%	20.2%	21.6%	23.4%
14%	14.1%	14.6%	15.3%	16.1%	17.2%	18.4%	19.9%	21.3%	23.2%	24.4%
16%	15.9%	16.5%	17.2%	18.0%	18.9%	20.0%	21.3%	22.7%	24.2%	25.4%
18%	18.1%	18.5%	19.1%	19.8%	20.7%	21.6%	22.9%	24.2%	25.8%	26.8%
20%	20.3%	20.5%	20.9%	21.6%	22.3%	23.4%	24.5%	25.8%	27.0%	28.4%
22%	22.1%	22.3%	23.1%	23.3%	24.3%	25.2%	25.9%	27.2%	28.4%	29.5%
24%	24.1%	24.4%	24.7%	25.4%	26.0%	26.9%	27.7%	28.9%	30.1%	31.3%
26%	26.2%	26.3%	26.6%	27.1%	27.9%	28.8%	29.9%	31.0%	32.0%	32.6%
28%	28.2%	28.2%	28.6%	29.3%	29.9%	30.3%	31.3%	32.6%	33.3%	34.7%
30%	29.6%	30.5%	30.6%	31.4%	31.6%	32.3%	33.2%	34.1%	35.0%	36.4%
32%	31.9%	32.1%	32.8%	32.9%	33.6%	34.1%	35.0%	36.3%	37.1%	38.0%
34%	34.1%	34.1%	34.6%	35.1%	35.3%	36.0%	36.7%	37.6%	38.5%	39.7%
36%	35.8%	36.6%	36.4%	37.0%	37.6%	38.2%	38.9%	39.3%	40.3%	41.3%
38%	38.0%	38.4%	38.5%	39.0%	39.1%	40.0%	40.4%	41.6%	42.3%	42.9%
40%	40.0%	40.1%	40.8%	41.0%	41.0%	41.4%	42.5%	43.0%	43.9%	45.2%

Appendix H. Process Used to Incorporate the Effect of Sampling Error

To incorporate the effect of sampling error, the demand reduction estimation process was replicated 100 times using 100 different randomly drawn samples. Figure H-1 describes the process for calculation methods that use control groups. A random sample was drawn from the broader population and randomly split into two groups.[31] For the first group, Group A, the demand reductions are simulated for each curtailment hour and subtracted from the loads. As a result, the actual demand reduction and the true counterfactual were known, enabling a comparison of how close the impacts calculate with the control group were to the true demand reductions. The unperturbed electricity use of the second group, Group B, was used to infer what AC electricity use would have been without the control operation and to estimate the demand reduction. The impacts were calculated in two ways, with a simple comparison of means and using a weather-matched difference-in-differences calculation. Both calculation methods are detailed in Section 4.5.4.

The demand reduction estimation process was also replicated 100 times using 100 different randomly drawn samples for the impact estimation tables. Figure H-2 shows the general framework used. The approach was similar to the one used for simulating random assignment to curtailments, but differed in subtle ways. First, a random sample of AC units is drawn from the population AC units with unperturbed end-use data.[32] Customers in the sample are used to develop the impact tables – based on weather condition bins, hour of day, climate region and the historical percent demand reductions. On the other hand, demand reductions were simulated for the entire population of AC units using the process described in Section 4 and Figure 4-1. Next, the tables based on the sample were used to provide an estimate of the demand reduction for each simulated curtailment event. In order to assess accuracy, the reductions predicted by tables were compared to the known impacts for the population. The process was repeated 100 times for each sample size to quantify the effect of sampling error.

[31] For the analysis, data from roughly 6,000 residential accounts was used and sampling for the bootstrap was conducted without replacement. For AC end-use data however, data from only 537 AC units were available, as a result the sampling was done with replacements; meaning that a particular AC unit could be sampled more than once.

[32] For AC end-use data however, data from only 537 AC units were available. Each of these units were replicated eight times to make up the population of AC units with end-use data. The samples to develop the matrix were randomly drawn from the 4,300 available observations. This approach is similar to bootstrapping, a standard method for incorporating sampling error.

Figure H-1: Incorporation of Sampling Error into Settlement Alternatives with Random Assignment

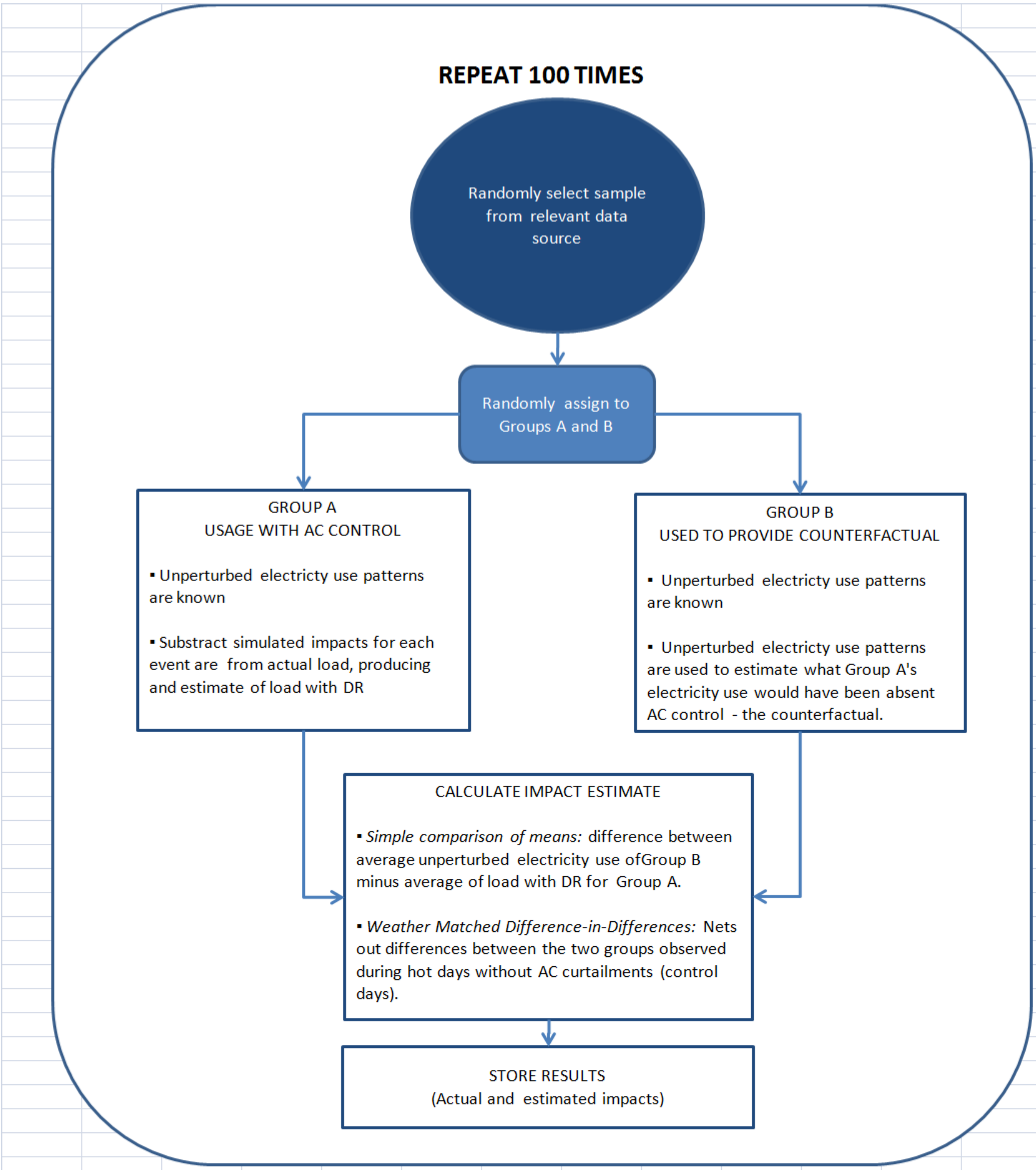
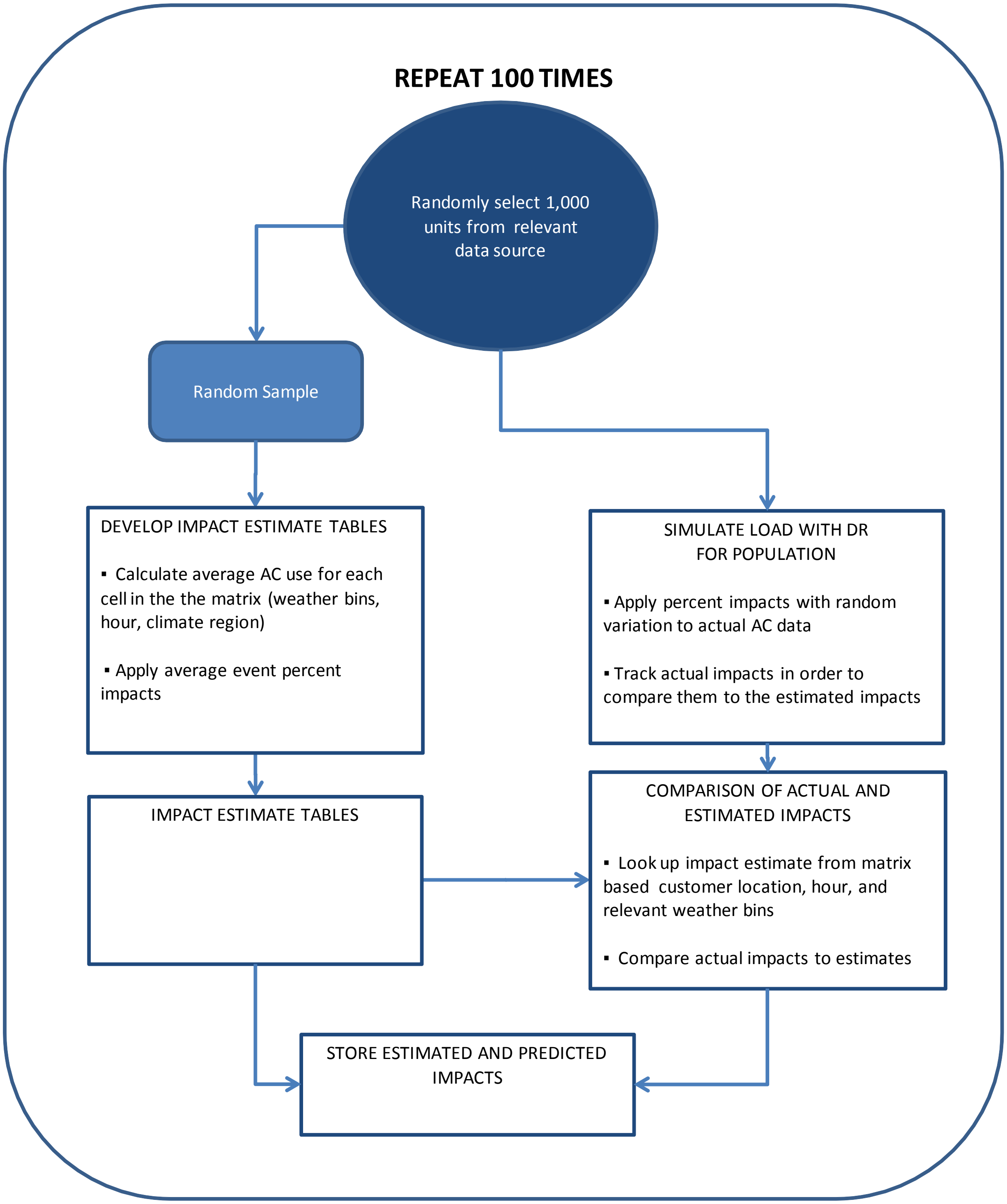


Figure H-2: Incorporation of Sampling Error into Impact Estimate Tables

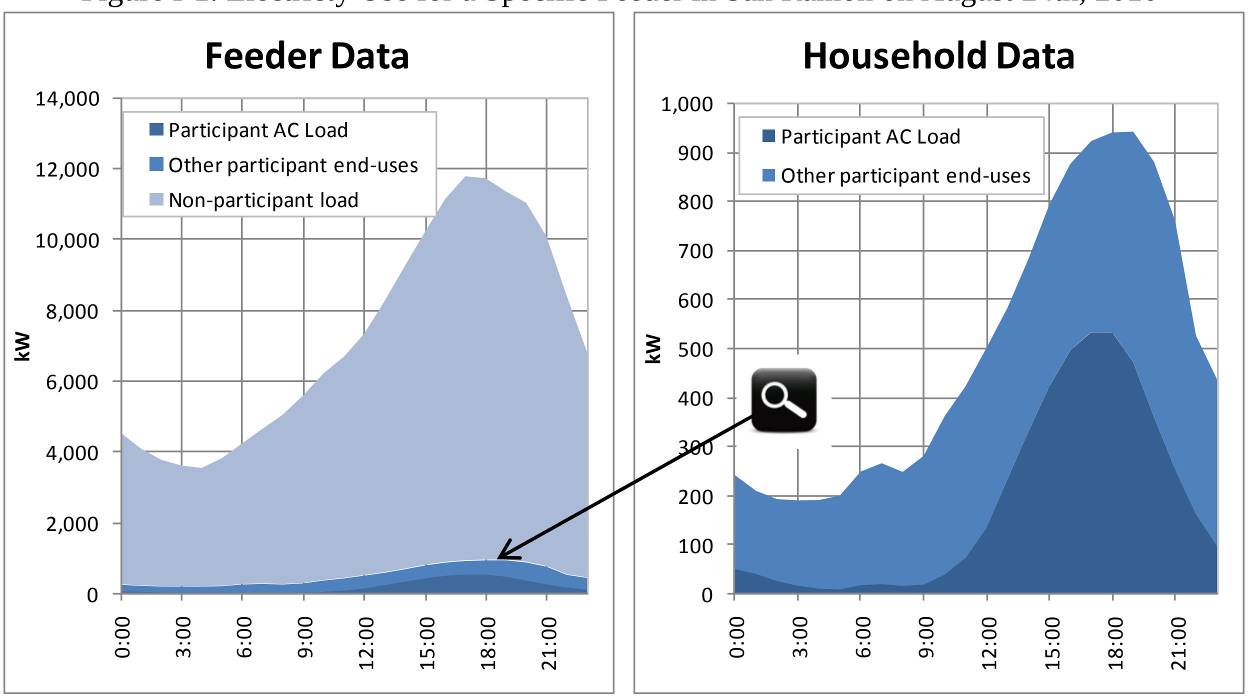


Appendix I. How the Data Source Affects Measurement

The data source has a strong influence on the ability to accurately detect and attribute changes in energy consumptions. The underlying change in electricity use is in the AC electricity use of customers enrolled in the program. Electricity use from other end-uses in participant households or from customers that are not in the program add to the background noise. The more extra data that is layered on top, the harder it is to detect and properly attribute changes in electricity use due to load control.

To illustrate, Figure I-1 shows electricity use for a specific feeder in one of the warmest cities in the Bay Area, San Ramon, on August 24th, 2010, a day when peak temperatures reached 103°F. The feeder had roughly 2,700 electric accounts at the time, some of which were commercial. It also had a relatively high penetration of the AC control program, SmartAC. The feeder had 266 electric accounts and 292 AC units under control at the time; which were roughly 10% of the accounts. The penetration exceeded the penetration in over 90% of the feeders in PG&E territory.

Figure I-1: Electricty Use for a Specific Feeder in San Ramon on August 24th, 2010



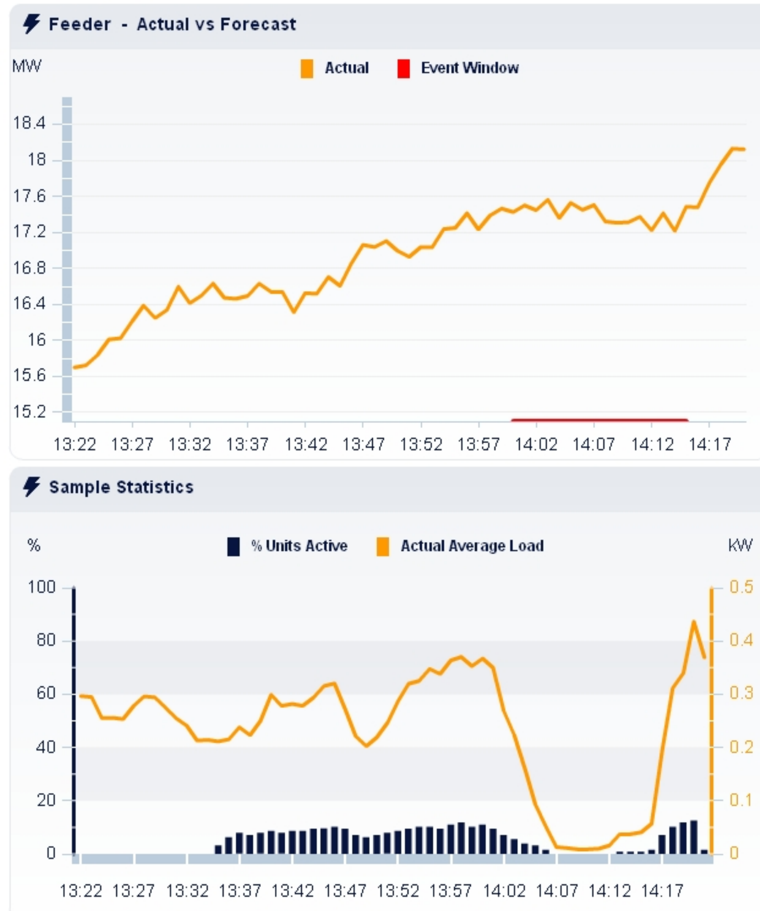
The graph on the left shows the electricity use on that feeder, while the graph on the right provides a more detailed look at the household and estimated AC use for customers enrolled in the AC control program. The overwhelming share of electricity use on the feeder, 92%, comes from electric accounts that are <u>not</u> enrolled in the program. In addition, for most hours of the day, the AC load is less than half of the household load. In any other, cooler day, the AC load would be a smaller share of the household electricity use. If we assume a 35% reduction of the AC electricity use during the hours with the highest AC use, at the feeder level it is necessary to distinguish a 175 kW (35% x 500 kW) change in electricity use from over 11,000 kW of feeder load. Because the change is relatively small, 1.5%, it is difficult to detect it and confidently eliminate alternative explanations, including random variation. In contrast, with household data of participants, the change is easier to observe and detect because it contains less noise. Assuming a 35% reduction in AC electricity use, it is necessary to distinguish 175 kW from approximately 900 kW, a far easier task. However, this is a best case scenario because the AC load is a smaller share of household load in most hours and days, which are typically cooler. Clearly, it is easiest to detect changes in AC electricity use when it is directly measured.

In practice, it is possible to increase penetration feeders so the electricity reductions are a larger share of the feeder load and can be observed. Employing more aggressive load control strategies such as 100% cycling or load shed also leads to larger reductions that are more easily detected. Both PG&E and SCE tested the ability to increase feeder penetration level and observe impacts with feeders versus with AC cycling. In both cases, the feeders were atypical in the degree of AC control saturation and the feeders selected. Even for highly saturated feeders, the results are mixed and depend on the date and time of operations.

Figure I-2 compares impacts on feeder loads to impacts when AC units are directly measured. It illustrates why impacts are more difficult to detect with feeder data, even in feeders with extremely high saturation of AC control. The example is from the PG&E 2009 Ancillary Services Pilot where 4 feeders were highly saturated and 2,000 AC compressors on 4 feeders were instructed to fully shut down 70 times for 15 minutes. The graph presents two different days, August 20th and September 8th, 2009, for the same feeders and AC units. A few observations are noteworthy. First, the reduction in AC load is easy to detect visually in both days, while it is more difficult to observe the impact using feeder data for September 8th. Second, although the impacts for August 20th are clearly visible of the feeder data, this is in part a product of the graph scale, which shows the feeder load above 20 MW, and the time resolution, which focuses on a single hour. Without the adjusted scales, the impacts on the feeder electricity load are less evident. Third, in both days, the impacts would be half or less if a 50% cycling strategy had been employed instead.

During the August 20th event window, the aggregate reduction in electricity use was 2.2 MW and the aggregated feeder electricity loads were 26 MW. The AC control produced a 8.5% reduction in feeder electricity loads. In contrast, the same reduction amounted to 76% of AC electricity use. For September 8th, the aggregate reduction in electricity use was 0.7 MW and the feeder loads were roughly 17 MW. The percent change in feeder electricity loads is smaller, 4.1%, and much harder to detect. In fact, a reader that was not informed of the event window could easily conclude the event occurred earlier, between 1:30 and 1:45 PM.

Figure I-2: Comparisons of Direct Load Control Impacts on Feeders and Air Conditioner Load

August 20th, 2009	September 8th, 2009
	

While it is possible to observe AC impacts on distribution feeder electricity data, in practice, few feeders will have high enough AC control penetration rates to be able to confidently and accurately detect changes in electricity use using that data source. Feeder data includes load variation from residential and commercial accounts, accounts with and without AC and multiple end-uses, many of which are unrelated to air conditioning. For most feeders, controllable AC load is too often too small compared to other electricity use on the feeder. To see impacts at the feeder level, the right feeder at the right hour during the right temperature conditions need to be selected and an aggressive control strategy needs to be employed. Even for feeders with extremely high AC control saturation levels, it can be hard to detect impacts using feeder data depending on weather conditions and hour of day. In testing the accuracy of settlement alternatives, rather than hand-pick highly saturated feeders, we employed a random sample of feeders to better understand the feasibility of feeder data for settlement on a program-wide basis.

Table I-1 summarizes the distribution circuit feeders in PG&E territory with customers enrolled in their AC load control program. There are approximately 2,800 distribution circuit feeders in PG&E territory and nearly 2,000 had customers enrolled in the AC control program when the data was extracted. The tables and the analysis exclude 66 feeders with less than 100 total population accounts since the results for these feeders are less reliable.

Table I-1: Characteristics of PG&E Feeders with Accounts in Air Conditioner Control

Penetration Deciles	Feeders	Central AC Saturation	Avg. Residential Feeder Accounts	Avg. Monthly kWh	Avg. Number of Accounts in SmartAC	SmartAC Participant Avg. Monthly kWh	% Penetration of Accounts
Top 10th	193	75.3%	1,980.7	801.5	199.7	766.4	10.0%
2nd	193	68.9%	2,002.7	726.6	143.7	734.8	7.1%
3rd	193	62.9%	1,795.2	713.6	102.7	746.6	5.7%
4th	193	57.8%	1,502.5	686.5	63.5	749.1	4.2%
5th	194	52.7%	1,640.9	693.7	47.6	748.2	2.9%
6th	193	49.2%	1,804.6	691.1	36.6	735.1	2.0%
7th	192	40.3%	1,821.9	618.1	24.9	663.5	1.4%
8th	194	31.7%	1,892.9	570.6	15.1	644.9	0.8%
9th	193	22.7%	1,700.9	537.1	6.0	650.7	0.3%
Bottom 10th	194	11.0%	2,570.1	462.1	1.8	576.6	0.1%
TOTAL	**1932**	**47.2%**	**1,871.5**	**650.0**	**64.1**	**707.2**	**3.5%**

In Table I-1, the top 10% of feeders with the highest penetration are grouped together, followed by the next 10% and so on. For each decile, the demographics of the average feeder are shown, including the total number of accounts, estimated central AC saturation and number of AC devices enrolled in the program. Not surprisingly, the AC control feeder penetration is higher when central AC saturation is higher. However, as the example of the San Ramon feeder demonstrated, even when over 10% of the accounts on a feeder are enrolled, the impacts of curtailment operations are a relatively small share of the feeder loads in extremely hot days and make up an even smaller share of the feeder loads in cooler days.

References

Bode, J., M. Perry, and S. Morgan. 2010. *Assessment of Settlement Baseline Methods for Ontario Power Authority's Commercial & Industrial Event Based Demand Response Programs.* Prepared for Ontario Power Authority.

Braithwait, S., D. Armstrong. 2009. *Load Impact Evaluation of California Statewide Aggregator Demand Response Programs. Volume 2: Baseline Analysis of AMP Aggregator Demand Response Program.* Prepared for the California Demand Response Measurement & Evalution Committee.

Christensen Associates Energy Consulting. 2010. *Highly Volatile-Load Customer Study.* Prepared for Southern California Edison, Pacific Gas and Electric, and San Diego Gas and Electric.

Coughlin K., M.A. Piette, C. Goldman, and S. Kiliccote. 2008. *Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California*. Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-63728. http://eetd.lbl.gov/ea/EMS/EMS_pubs.html

Eto, J., Nelson-Hoffman, C. Torres, S. Hirth, B. Yinger, J. Kueck, B. Kirby (Oak Ridge National Laboratory), C. Bernier, R. Wright, A. Barat, D. Watson. 2007. *Demand Response Spinning Reserve Demonstration.* Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-62761.

Eto, J., C. Bernier, P. Young, D. Sheehan, J. Kueck, B. Kirby, J. Nelson-Hoffman, and E. Parker (2009). *Demand Response Spinning Reserve Demonstration — Phase 2 Findings from the Summer of 2008.* Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-2490E.

Gifford, W., S. Bodmann, P. Young, J. Eto, J. Laundergan. 2010. *Customer Impact Evaluation for the 2009 Southern California Edison Participating Load Pilot.* Lawrence Berkeley National Laboratory. LBNL-3550E.

KEMA, Inc. 2003. *Protocol Development For Demand Response Calculation—Findings and Recommendations.* Prepared for the California Energy Commission. http://www.energy.ca.gov/reports/2003-03-10_400-02-017F.PDF

KEMA, Inc. 2011. *PJM Empirical Analysis of Demand Response Baseline Methods.* Prepared for the PJM Markets Implementation Committee. http://pjm.com/markets-and-operations/demand-response/~/media/markets-ops/dsr/pjm-analysis-of-dr-baseline-methods-full-report.ashx

Kirby, B. 2003. *Spinning Reserve From Responsive Loads.* Oak Ridge National Laboratory. ORNL/TM-2003/19.

Kirby, B. 2006. *Demand Response For Power System Reliability: FAQ.* Oak Ridge National Laboratory. ORNL/TM-2006/565.

Kueck, J., B. Kirby, R. Staunton, J. Eto, C. Marnay, C. Goldman, C.A. Martinez. 2001. *Load As a Reliability Resource in Restructured Electricity Markets.* Prepared for the U.S Department of Energy. http://certs.lbl.gov/pdf/load-reliability.pdf

Kueck K., B. Kirby, M. Ally, and. Rice. 2009. *Using Air Conditioning Load Response for Spinning Reserve.* Oak Ridge National Laboratory. ORNL/TM-2008/227.

Mathieu, J., D.S. Callaway, and S. Kiliccote. 2011. *Variability in Automated Responses of Commercial Buildings and Industrial Facilities to Dynamic Electricity Prices.* Energy and Buildings. *Volume 43, Issue 12, December 2011*

Quantum Consulting and Summit Blue Consulting. 2004. *Working Group 2 Demand Response Program Evaluation – Program Year 2004 Final Report*. Prepared for the California Public Utilities Commission.

Sullivan, M., J. Bode, P. Mangasarian. 2009. *2009 Pacific Gas and Electric Company SmartAC Ancillary Services Pilot.* Prepared for Pacific Gas and Electric.